AN INVESTIGATION OF THE EFFECTS OF OPPORTUNITIES TO RESPOND AND INTELLIGENCE ON SIGHT WORD RETENTION USING INCREMENTAL REHEARSAL

by

Kade Ryan Johnson

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Educational Psychology

The University of Utah

August 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Kade Ryan Johnson**

has been approved by the following supervisory committee members:

Daniel E. Olympia	, Chair	**5-1-2012** Date Approved
Lora Tuesday-Heathfield	, Member	**5-1-2012** Date Approved
William R. Jenson	, Member	**5-1-2012** Date Approved
Robert E. O'Neill	, Member	**5-1-2012** Date Approved
Robert J. Volpe	, Member	**5-1-2012** Date Approved

and by **Elaine Clark** , Chair of

the Department of **Educational Psychology**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

High opportunities to respond (OTR) have been touted as being a key factor in a popular and effective drill procedure called incremental rehearsal (IR). However, IR has also been criticized because it takes more instructional time than other drill procedures and can be less time efficient. The current study compared the effectiveness and efficiency of a high (44 OTR), medium (27 OTR), and low (14 OTR) OTR version of IR using 23 sixth, seventh, and eighth grade students. Eleven students had low IQ scores (M = 68.18, SD = 6.82) and 12 students had average IQ scores (M = 101.00, SD = 6.63). Students were taught six Esperanto words during each condition. Effectiveness was based on 1- and 3-week retention measures and efficiency was determined by evaluating the number of words initially learned, retained at 1 week, and retained at 3 weeks per minute of instruction. A within-subjects repeated measures ANOVA was used to evaluate both the number of words recalled and the number of words recalled per minute of instruction. Results for both participant groups demonstrated that the high and medium OTR conditions were equally effective. However, students retained the most words per minute of instruction during the medium OTR condition, so this condition was the most time efficient. The number of words retained 1 week after instruction during the high and medium OTR conditions was not significantly different for the two participant groups. However, the number of words retained 3 weeks after instruction during the

high and medium OTR conditions was significantly different for the two participant groups, suggesting that students in the low IQ group forgot a significant number of words between the 1- and 3-week retention measures, whereas the average IQ group did not. Limitations and implications for practitioners and future researchers are discussed.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND REVIEW OF THE LITERATURE

Introduction

Sight word recognition, particularly for students with disabilities, is a fundamental reading skill that often predisposes reading fluency and comprehension (Browder & Xin, 1998; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003). Drill and practice methods typically involve the use of flash cards and are frequently used to teach sight words. Commonly used drill and practice methods include traditional drill and practice, interspersal methods, and incremental rehearsal (IR). Known words are interspersed and practiced among unknown target words during interspersal and IR procedures, whereas only unknown target words are presented during traditional drill and practice. Sufficient repetition and practice opportunities are key aspects of drill and practice procedures and are particularly fundamental to reading instruction for individuals with below average intelligence (Allor, Champlin, Gifford, & Mathes, 2010).

IR is a popular type of interspersal procedure in which unknown words are systematically interspersed and practiced among known words, typically using a ratio of 90% known words to 10% unknown words (MacQuarrie, Tucker, Burns, & Hartman, 2002; Tucker, 1989). As an unknown word reaches the mastery criterion, it is folded-in and continually practiced as a known word, thus creating additional practice opportunities (Shapiro, 2011). The systematic folding-in procedure used in IR sets it apart from other

interspersal procedures. Several studies have demonstrated the effectiveness of IR in comparison to traditional drill and practice, as well as other types of interspersal procedures (Burns, Dean, & Foley, 2004; MacQuarrie et al., 2002; Nist & Joseph, 2008). Despite its effectiveness, however, IR has been criticized because it takes more time than other drill and practice procedures and is less efficient (Nist & Joseph, 2008; Skinner, 2008). Efficiency, or learning rate, is traditionally determined by dividing the number of words learned by the number of instructional minutes (Skinner, Belfoire, & Watson, 1995). However, when the efficiency of IR was calculated using the number of words retained, rather than the number of words learned, IR was found to be equally, or more efficient than other drill methods (Burns & Boice, 2009; Burns & Sterling-Turner, 2010). High opportunities to respond (OTR; Greenwood, Delquadri, & Hall, 1984) are touted as being a key factor of IR (Burns, 2007a; Szadokierski & Burns, 2008). However, due to the varying OTR to each unknown word in the IR procedure, systematic and accurate measures of the effectiveness and efficiency of IR are limited within the research (Volpe, Mulé, Briesch, Joseph, & Burns, 2011). Likewise, when using IR, the optimal OTR necessary for retention, particularly for individuals with below average intelligence, has received only minimal attention, and is in need of future investigation (Szadokierski & Burns, 2008). Thus, an examination of the optimal OTR necessary to ensure retention of words taught using IR is warranted. Additionally, since IR appears to negate the effects of intelligence on word retention (Burns & Boice, 2009), correlations between intelligence and optimal levels of OTR should also be examined.

This literature review begins by outlining key components of reading and word recognition, specifically emphasizing necessary components for students with intellectual

disabilities. Commonly used methods to teach sight words, such as traditional drill and practice, interspersal procedures, and IR are then outlined. As related to interspersal procedures, the concepts of efficiency, drill ratio, and OTR are introduced and discussed. A comprehensive description of IR is provided thereafter, followed by several examples of IR in the research. The remainder and bulk of this review is devoted to an in-depth analysis of OTR within the IR procedure. This part is organized into six major sections. The first section begins by discussing the varying OTR to unknown words when using IR. Second, several methodological oversights regarding how OTR have been measured and conceptualized in previous research is discussed. Next, two studies illustrating the importance of precisely measuring OTR to each unknown word are outlined. In the fourth section, the hypotheses of several researchers regarding possible causal mechanisms of IR, in addition to providing high levels of OTR, are explored. Fifth, although the research examining optimal levels of OTR is minimal, the extant literature is reviewed. Finally, a discussion of intelligence, as it relates to IR and OTR, is provided. This review concludes by summarizing the critical aspects of previous research that have guided the ensuing study. Based on this summary, a problem statement and study rationale are introduced, followed by specific research questions.

Word Recognition

Gravois and Gickling (2008) describe reading as an orchestration between various skills in order to gain meaning from text. Word recognition is a basic feature of efficient reading that leads to reading fluency and subsequent reading comprehension (Burns et al., 2002; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Levy, Abello, & Lysynchuk, 1997; Schwanenflugel, Meisinger, Kuhn,

Strauss, & Morris, 2006). Word recognition is defined as "the ability to identify, pronounce, and know the meaning of words" (Gravois & Gickling, 2008, p. 502). Research indicates that students with intellectual disabilities can effectively learn isolated word reading skills when appropriate instructional interventions are used (Browder, Wakeman, Spooner, Ahlgrim-Delzell, & Algozzine, 2006; Browder & Xin, 1998; Burns, 2007a; Burns & Boice, 2009; Joseph & Seery, 2004). Although an increasing amount of research demonstrates effective phonics approaches for teaching words to children with intellectual disabilities (Conners, 1992; Joseph & Seery 2004), the majority of the research on word reading for this population is related to sight word acquisition (Browder & Xin, 1998).

Ehri (2005) identified sight word reading as the "most efficient, unobtrusive way to read words in text" (p. 167). This is important for children with intellectual disabilities because word recognition is not only a precursor to fluent reading and comprehension in the classroom, but is a necessary life skill within the community (Schloss, Alper, & Young, 1995). Browder and Xin (1998) conducted a meta-analytic review of 48 studies that taught sight words to individuals with mild to severe disabilities. Based on this review, sight word instructional methods led to rapid acquisition of skills. Percentage of nonoverlapping data (PND) was calculated to evaluate the overall treatment effectiveness of sight word approaches, as well the effectiveness of sight word approaches for individuals with varying ages and degrees of disability. Across all sight word interventions, a median PND of 91 (range 63-100) demonstrated sizeable nonoverlaping of data between baseline and treatment phases, suggesting a significant treatment effect. Although sight word reading approaches were shown to be effective across varying

degrees of disability, statistically significant differences ($p<.01$) were found between those with mild intellectual disabilities and those with severe intellectual disabilities. There were no significant differences found between adults, secondary students, and elementary students. These findings support the effectiveness of sight word instructional approaches for individuals with intellectual disabilities.

Common Sight Word Instructional Approaches

Drill and Practice Approaches

In their seminal paper originally published in 1929, Symonds and Chase (1992) posit that coupled with the right motivation, "the most effective device that can be applied to learning is to increase the amount of drill or practice (p. 289)." Haring and Eaton (1978) described drill as a procedure that uses repetition of responses for the purpose of learning. Alternatively, they described practice as a problem solving approach using learned responses. Academic remediation via drill and practice techniques allows for high repetition of new items during early learning (Daly, Hintze, & Hamler, 2000; Roberts & Shapiro, 1996) which in turn leads to the development of skill fluency and automaticity (Cooke, Guzaukas, Pressley, & Kerr, 1993). Drill and practice tasks represent an important remedial step to teaching the basic foundational academic skills necessary for higher order academic skills (Burns, 2004b). For individuals with intellectual disabilities, this is particularly important. In their review of methods to improve reading instruction for students with intellectual disabilities, Allor, Champlin, Gifford, and Mathes (2010) concluded that sufficient repetition and opportunities to practice are key aspects of reading instruction for this population.

Traditional Drill and Practice

A variety of drill and practice methods have been effectively utilized in educational research and practice (MacQuarrie et al., 2002; Symonds & Chase, 1992). A commonly used method is called traditional drill and practice. This is a type of flashcard technique in which unknown stimuli (e.g., sight words, math facts, etc.) are presented one after the other until a predetermined mastery criterion is met for each stimulus (Tan & Nicholson, 1997). Flashcards are initially presented in conjunction with instructor modeling and practice. Corrective feedback is provided if necessary. Traditional drill and practice flashcard approaches have been shown to increase the number of words read in isolation, and are responsible for improved reading fluency and comprehension (Burns et al., 2004; Joseph & Schisler, 2007; Tan & Nicholson, 1997).

Interspersal Procedures

Unlike traditional drill and practice, interspersal procedures involve the systematic alteration of trials to include known or previously mastered tasks or stimuli in combination with unknown tasks or stimuli (Neef, Iwata & Page, 1980). The use of interspersal techniques has demonstrated increased acquisition of material (Dunlap & Koegel, 1980), rates of learning (Cooke et al., 1993), and task completion (Skinner, 2002). Neef, Iwata and Page (1977; 1980) compared the effectiveness of an interspersal procedure and a control condition on the acquisition and retention of spelling words for children described as mentally retarded. For the interspersal condition, 10 known spelling words and 10 unknown spelling words were interspersed and practiced with students. In the control condition, 10 unknown spelling words were practiced using a traditional drill and practice approach, without interspersing known words. Participants demonstrated

better acquisition and retention of new spelling words during the interspersal condition than they did during the control condition. These two studies demonstrated the additive benefits of interspersing known material among unknown material during drill and practice procedures.

Browder and Shear (1996) investigated the effectiveness of an interspersal treatment package to teach unknown sight words to three middle school students with moderate mental retardation and behavior disorders. Students were taught 10 weather related sight words, interspersed among 10 previously known sight words. Dependent variables included the number of unknown sight words correctly read during a test, the percentage of error responses for unknown words, and generalization of the learned words. Using the interspersal procedure, participants learned each of the 10 sight words within 2 to 6 weeks. Previously, the students had learned an average of only 30 total sight words throughout the previous several school years using a time-delay procedure. This demonstrates the relatively quick acquisition of sight words for students with intellectual disabilities when using interspersal procedures.

The positive effect of interspersing known items with unknown items has been consistently demonstrated throughout the literature (Cuvo, Davis, & Gluck, 1991; Dunlap, 1984; Roberts & Shapiro, 1996; Roberts, Turco, & Shapiro, 1991; Skinner, Fletcher, Wildmon, & Belfiore, 1996; Skinner et al., 1999). In addition to the academic gains resulting from interspersal procedures, social validity data demonstrate favorable acceptability (Cates & Skinner, 2000; Cates et al., 1999; Skinner, Robinson, Johns, Logan, & Belfiore, 1996) as well as improved likelihood of academic engagement (McCurdy, Skinner, Grantham, Watson, & Hindman, 2001). Students perceive tasks to be

less demanding when known content is included and are more likely to remain engaged and complete tasks (Dunlap, 1984; Neef et al., 1980).

Efficiency of interspersal procedures. While interspersing known material with unknown material is an effective instructional strategy (Burns et al., 2004; MacQuarrie et al., 2002), some researchers argue that this process increases the amount of time the intervention takes and leads to delayed learning rates (Cates et al., 2003; Cooke et al., 1993; Joseph & Nist, 2006; Nist & Joseph, 2008; Schmidgall & Joseph, 2007; Skinner, 2008). Learning rate is determined by dividing the number of items learned by the amount of instructional time (e.g., number of minutes; Skinner, Belfiore, & Watson, 1995). This calculation yields a measure of efficiency, which can be compared across two or more interventions. The importance of intervention efficiency is highlighted in the following hypothetical example: If a student learns five sight words using intervention A, and also learns five sight words using intervention B, at face value it would appear that these two interventions are equally effective and suitable for use with students. However, if intervention A takes 10 minutes to implement and intervention B takes only 5 minutes to implement, one can clearly see the benefit of using intervention B because it takes half the time as intervention A, but yields the same result. Thus, in order to maximize instructional time and student learning, intervention efficiency should be considered when selecting, implementing, and evaluating interventions (Skinner, 2008; Skinner, 2010; Skinner, Belfiore, Mace, Williams-Wilson, & Johns, 1997).

To illustrate the importance of considering instructional time and student learning rates, Cates et al. (2003) compared the effectiveness and efficiency of three spelling instructional methods. Five second-grade general education students identified as

struggling spellers participated in a traditional drill and practice condition, an interspersal condition, and a high-p sequencing condition. For the traditional drill and practice condition, only unknown spelling words were presented to the students. During the interspersal condition, three known spelling words were presented following every third unknown spelling word. During the high-p-sequencing condition, six unknown spelling words were interspersed and rehearsed with 18 known spelling words by presenting three known words prior to each unknown word. This procedure was based on the concept of behavioral momentum suggesting that the probability of an individual engaging in a challenging or unfavorable task is higher if it is precluded by less challenging or more favorable tasks (Nevin, Mandell, & Atak, 1983).

Using an alternating treatments design, instructional effectiveness and efficiency were measured for each participant across each condition. Instructional effectiveness was determined by measuring the total number of words learned during each condition. As a measure of efficiency, the number of words mastered during each session was multiplied by 60 seconds then divided by the instructional time in seconds. Thus, the student's learning rate, or number of words learned per minute of instruction, was compared across conditions to determine which condition was the most time efficient. Results indicated that students learned about the same number of words during each condition, suggesting equal effectiveness for each spelling intervention. However, when instructional time was considered, learning rates were highest in the traditional drill and practice condition, followed by the interspersal condition, and finally the high-p sequencing condition. Thus, although students successfully learned new spelling words during each condition, traditional drill and practice was determined to be the most efficient approach. These

results demonstrate the importance of considering learning rates and intervention efficiency, not just intervention effectiveness, when selecting and implementing academic interventions.

Drill ratio within interspersal procedures. First introduced by Betts (1946), and later refined by Gickling and Armstrong (1978), the concept of teaching within the optimal instructional range is a widely accepted component of effective instructional practices (Gravois & Gickling, 2008). The goal of interspersal procedures is to create a ratio of known to unknown tasks that are not too difficult or too simple in order to establish an optimal instructional range. Gickling and Armstrong coined the terms "frustrational," "instructional," and "independent" to describe the difficulty of reading passages. Research has consistently supported the assertion that reading passages should ideally include 93% to 97% known words in order to be within the instructional range (Gickling & Armstrong, 1978; Gickling & Rosenfield, 1995; Shapiro, 2011; Shapiro & Ager, 1992). However, the recommended percentage of known words to include in drill and practice methods has greatly varied in the research (Burns, 2004b). Early researchers evaluated interspersal drill and practice methods with ratios ranging from less than 50% to 90% known material (Gickling & Thompson, 1985; Neef et al., 1980; Roberts, Turco, & Shapiro 1991; Robinson & Skinner, 2002; Tucker, 1989), with each ratio demonstrating some positive effect. In their review of interspersal research on drill and practice methods, Cooke et al., (1993) concluded that although the beneficial effects of interspersing known items among unknown items during drill and practice procedures was apparent, the optimal ratio of known to unknown items had not been clearly determined.

In an effort to bring some clarity to this matter, Burns (2004b) conducted a meta-analysis to evaluate the effect of drill ratio on the retention of material taught via interspersal procedures. Studies using four levels of known material were evaluated: less than 50% known, 50% to 60% known, 70% to 85% known, and 90% known. Overall findings indicated that drill ratios with higher percentages of known material were more effective than those with smaller ratios. Drill ratios using 70% to 85% known material, as well as those using 90% known material demonstrated comparable mean effect sizes with coefficients of 1.22 (*SD* = .44) and 1.19 (*SD* = .37), respectively. Drill ratios using 50% to 60% known material also yielded a large, though somewhat smaller, mean effect size with a coefficient of .85 (*SD* = .65). Drill ratios using 50% or less known material demonstrated a small to medium mean effect size with a coefficient of .49 (*SD* = .49). Based on these data, ratios of 50% known material or higher led to large effects. However, a precise or optimal drill ratio for drill and practice methods did not emerge from this analysis. Additionally, the analysis by Burns (2004b) did not consider the amount of time each intervention took under each drill ratio. Thus, intervention efficiency was not considered.

Opportunities to respond (OTR) within interspersal procedures. Regardless of the drill ratio used, the process of integrating unknown material with known material creates fewer opportunities to respond (OTR) to unknown material within a given instructional period (Cates et al., 2003). This is an important consideration because previous researchers have hypothesized that a high level of OTR to unknown content, rather than the ratio of known to unknown content, may be responsible for increased retention of material taught via interspersal drill procedures (MacQuarrie et al., 2002).

This claim was evaluated by Szadokierski and Burns (2008) by examining the effect of varying drill ratios and OTR on the retention of unknown words taught using an interspersal procedure. Using a 2 x 2 analysis of variance (ANOVA) with repeated measures, a significant main effect was found for OTR, whereas nonsignificant effects were found for drill ratio and the interaction between OTR and drill ratio. Thus, OTR were significantly correlated with word retention, whereas the ratio of known to unknown words had little correlation with retention. Cohen's *d* effect sizes were also calculated. Increasing OTR from low to high yielded a large effect ($d = 2.46$), but increasing the number of known words from moderate to high yielded a small effect ($d = 0.16$). Thus, consideration to the number of OTR afforded by a drill ratio may be more critical than the drill ratio itself when using interspersal procedures.

Greenwood et al. (1984) describes OTR as the interaction between instructional stimuli and a student's successful response to those stimuli. With each additional exposure to stimuli, the relationship between the stimuli and response is strengthened, leading to more correct responses and fewer incorrect responses. As students are presented with increased practice opportunities during learning trials, learning is improved (Burns, 2004b). High levels of OTR have led to improved fluency (Skinner, Turco, Beatty, Rasavage, 1989), accuracy (Albers & Greer, 1991), retention (Logan & Klapp, 1991) and overall learning (Greenwood et al., 1984; Skinner, Fletcher, & Hennington, 1996). Additionally, high levels of OTR have been identified as a fundamental aspect of learning for individuals with disabilities (Chase & Symonds, 1992; Kamps, Dugan, Leonard, & Daoust, 1994), particularly for those with below average intelligence (Burns, 2007a; Burns & Boice, 2009). A more in depth analysis of OTR,

particularly within an interspersal procedure called incremental rehearsal, is included later in this review.

Incremental Rehearsal (IR)

A systematic interspersal procedure called incremental rehearsal (IR) has gained considerable attention in research and practice. IR is a flash card drill and practice technique that presents known and unknown material in a precise and carefully arranged sequence (Tucker, 1989). In the published research, IR has most frequently been used as an effective method of teaching sight words (Burns & Sterling-Turner, 2010; Burns et al., 2004; MacQuarrie et al., 2002; Nist & Joseph, 2008). However, it has also been used to teach letter identification skills (Bunn, Burns, Hoffman, & Newman, 2005), letter sounds (Volpe, Burns, DuBois, & Zaslofsky, 2011), and math facts (Burns, 2005; Coding, Archer, & Connell, 2010). IR has been effectively implemented across the age span, ranging from preschool students (Bunn et al., 2005) to adults with cognitive delays (Burns & Kimosh, 2005), as well as with various populations such as students with limited English proficiency (Matchett & Burns, 2009), learning disabilities (Burns, 2005; Burns et al., 2004; Burns & Boice, 2009), intellectual disabilities (Burns, 2007a; Burns & Boice, 2009), and at-risk general education students (Schmidgall & Joseph, 2007). IR is supported by key instructional components such as spaced repetition (Dempster, 1991; Glenberg & Smith, 1981), errorless learning (Browder & Shear, 1996), gradual introduction of material (Schnorr, 1989), teaching to automaticity (Dehaene & Akhavein, 1995; Jones & Christensen, 1999), high opportunities to respond (Burns, 2007a; Greenwood et al., 1984; MacQuarrie et al., 2002; Szadokierski & Burns, 2008) and a high percentage of known items (Burns, 2004b). Although the effectiveness and

foundational principles of IR have been clearly demonstrated in the research, some students prefer traditional drill and practice over IR because it takes less time (Nist & Joseph, 2008).

IR procedure. IR typically uses a ratio of 90% known to 10% unknown words (MacQuarrie et al., 2002; Tucker, 1989) and incrementally presents words using the following sequence: U1, K1, U1, K1, K2, U1, K1, K2, K3, U1, K1, K2, K3, K4, U1, K1, K2, K3, K4, K5, U1, K1, K2, K3, K4, K5, K6, U1, K1, K2, K3, K4, K5, K6, K7, U1, K1, K2, K3, K4, K5, K6, K7, K8, U1, K1, K2, K3, K4, K5, K6, K7, K8, K9. In this sequence, U1 is the first unknown word taught; K1 represents the first known word; K2 represents the second known word, and so on. Once this entire sequence has been rehearsed, rather than removing the first unknown word from the practice deck of flash cards, it is placed in the first known word position, and the ninth known word is removed from the sequence. The second unknown word is now introduced to the practice deck and is placed in the position previously occupied by the first unknown word. Thus, when the second unknown word is introduced, the new unknown word is placed in the U1 position, the previously practiced unknown word moves to the K1 position; the known word previously occupying the K1 position now occupies the K2 position, and so on. This procedure, described as "folding-in" by Shapiro (2011), is continued until the final unknown word has been presented. For example, if ten unknown words were taught during an IR session containing nine known words, at the end of that session, only the targeted unknown words would be found remaining in the practice deck because each of the original known words would have been replaced by folding-in the unknown words. By folding-in unknown words, each word is afforded multiple presentations, continued

practice, and high levels of OTR (Burns, 2007a; MacQuarrie et al., 2002; Szadokierski & Burns, 2008). The folding-in component is a unique aspect of IR and significantly varies from other interspersal techniques which typically remove known words from the practice deck once a predetermined mastery criterion is met.

Examples of IR in the research. Burns and Kimosh (2005) examined the effectiveness of IR by teaching sight words to individuals with moderate intellectual disabilities. They used a multiple baseline across participants design. Participants included two adult female students, ages 19 and 21, who attended a "young adult" special education program within the local school district. Both participants met special education eligibility criteria as students with Moderate Mental Retardation, with intelligence standard scores of 48 and 55, adaptive behavior standard scores ranging from 70 to 80, and academic achievement below the 6th percentile. During eight to ten sessions, one participant was taught 30 hygiene related sight words and the other participant was taught 30 shopping list sight words. Words were incrementally rehearsed, as explained previously, using a ratio of 90% known to 10 % unknown words. Students were taught the correct pronunciation, definition, and usage of each word in a sentence. The dependent variable was the number of words read correctly per minute on sight word fluency probes. During the baseline phase, Participant 1 correctly read an average of 14.25 words per minute and Participant 2 correctly read an average of 1 word per minute. During the treatment phase, words read correctly per minute increased to 27.38 and 10.63, respectively. Strong effect sizes and 100% PND were demonstrated for both participants. The same sight word fluency measures were used 3 weeks later and were consistent with posttreatment data, indicating good maintenance. Results of this study

suggest that IR is an effective method of teaching sight words to individuals with moderate cognitive delays.

MacQuarrie et al. (2002) evaluated the effectiveness of IR, Drill Sandwich (DS; a type of interspersal procedure), and traditional drill (TD) for teaching words from the Esperanto International Language using a within-group design. Each of the 25 third-grade and 26 seventh-grade participants were taught nine unknown Esperanto words using each drill method. Therefore, each student was taught a total of 27 words across three separate conditions. The Peabody Picture Vocabulary Test-Revised (Dunn, L. M. & Dunn, L. M., 1981) was administered to each participant prior to the intervention to ensure adequate receptive vocabulary. Only participants with standard scores of 80 or higher were included in the study. Across all participants, the mean PPVT-R standard score was 101.14 (*SD* = 11.69).

In the TD condition, the nine unknown Esperanto words were presented to participants until they were able correctly pronounce and state the English translation three times. The DS condition included three sets of three unknown Esperanto words. Each set also included six known words and were interspersed as follows: K1, K2, K3, U1, K4, K5, U2, K6, U3. During the IR condition, each of the nine unknown Esperanto words was incrementally folded-in to the drill sequence in the following order: U1, K1, U1, K1, K2, U1, K1, K2, K3, U1, K1, K2, K3, K4, U1, K1, K2, K3, K4, K5, U1, K1, K2, K3, K4, K5, K6, U1, K1, K2, K3, K4, K5, K6, K7, U1, K1, K2, K3, K4, K5, K6, K7, K8, U1, K1, K2, K3, K4, K5, K6, K7, K8, K9. As described previously, U1 was the first unknown word, K1 was the first known word, K2 was the second known word, and so on. Once rehearsed, unknown words were incrementally folded-in until all nine words

had been introduced and practiced. Error correction was provided during each condition if the student responded incorrectly. For each treatment condition, word retention was measured 1, 2, 3, 7, and 30 days after instruction.

Across each retention measure, the IR procedure led to the highest word retention, followed by TD then DS. Words taught during the IR condition were also maintained longer than words taught during the TD and DS conditions. From the 1st day retention measure to the 30th day retention measure, a decrease in words retained of only 9.2% was observed for the IR condition, while the DS and TD conditions resulted in a 32.8% and 35.2% decrease in words retained, respectively. Due to the folding-in component, the IR condition had the highest OTR, followed by TD then DS. Thus, the effectiveness of each condition was directly proportionate to the OTR within each condition. In consideration of the high level of OTR found in the IR condition, it is not surprising that the IR treatment sessions took about twice as long as the DS sessions.

Although MacQuarrie et al. (2002) made mention of the varying OTR and time required to complete each condition, no attempt was made to hold either of these variables constant in order to objectively compare the efficiency of each condition. Nist and Joseph (2008) conducted a study similar to MacQuarrie et al. by teaching sight words using IR, an interspersal procedure, and traditional drill and practice. This study differed from the study by MacQuarrie and colleagues because intervention efficiency was one of the main dependent variables. Due to the positive link between OTR and retention (Burns, 2007a, MacQuarrie et al., 2002), Nist and Joseph attempted to maintain a consistent level of OTR across conditions in order to quantify and compare intervention efficiency. Thus, six unknown English sight words were presented a total of nine times

across each treatment condition. Using a counterbalanced alternating treatment design, six first-grade students experienced all three conditions during each session. Thus, students were taught a total of 18 unknown words per session. The IR condition was implemented as described previously by incrementally interspersing the unknown words among nine known words using a drill ratio of 90% known to 10% unknown words. The interspersal procedure, using a ratio of 33% known to 67% unknown words, consisted of a known word being presented prior to every third unknown word in this manner: K1, U1, U2, U3, K2, U4, U5, U6, K3. The traditional drill and practice procedure consisted only of the six unknown words. As dependent variables, instructional effectiveness was evaluated by examining the cumulative number of words read correctly on next-day retention measures and instructional efficiency was evaluated by examining the cumulative rate of words retained on next-day retention measures per minute of instruction.

Regarding intervention effectiveness, results were consistent with MacQuarrie et al. (2002) with next-day word retention being highest in the IR condition. Regarding intervention efficiency, based on next-day retention measures, students learned the fewest words per minute of instructional time using IR. Traditional drill and practice was determined to be the most efficient approach, followed by the interspersal procedure, and finally IR. These results are consistent with previous studies that also suggest that more words are learned per minute of instruction using TD than with interspersal procedures. (Cates et al., 2003; Joseph & Nist, 2006; Joseph & Schisler, 2007; Schmidgall & Joseph, 2007). However, as pointed out by Nist and Joseph (2008), variables such as word maintenance over time, as well as generalization of skills, should also be considered

when evaluating intervention efficiency. Word maintenance and generalization were measured 5 and 6 days, respectively, after the final instructional session. Results indicated word maintenance and generalization were highest for IR, followed by the interspersal procedure, and finally traditional drill and practice. These results are promising and further suggest the effectiveness of IR, particularly relative to the procedure's longitudinal effects. However, the additional time IR takes may be a concern in some situations, leading some practitioners to select more time efficient interventions. However, since efficiency in this study was calculated using next-day retention measures, it is unknown if traditional drill and practice would have remained the most efficient intervention if it had been calculated based on the maintenance measures administered 5 or more days post-intervention.

In a recent study by Burns and Sterling-Turner (2010), 25 fourth-grade general education students were taught six unknown Esperanto words using IR, and six unknown Esperanto words using traditional drill (TD). In contrast to the study by Nist and Joseph (2008), the efficiency of each instructional method was calculated using both the initial learning data, as well as data from 1-week retention measures. TD was implemented as described in previous studies. IR was also implemented as described in previous studies except eight, rather than nine known words were used. This was done in order to save time, since no effectual differences were previously found when using eight known words as opposed to nine known words (Burns, 2004b). For the initial learning efficiency calculation, the number of words taught was divided by the number of instructional minutes within each condition. For the 1-week retention efficiency calculation, the number of words maintained was divided by the number of instructional minutes within

each condition. These two calculations yielded a measure of efficiency to show the number of words learned and maintained per minute of instruction.

Across participants, the TD condition took an average of 3.52 minutes to complete and the IR condition took an average of 8.90 minutes to complete. This equated to 1.78 words initially learned per minute using TD and 0.69 words initially learned per minute using IR. These results are similar to those found by Nist and Joseph (2008) when efficiency was calculated using next-day retention measures. However, when efficiency was calculated using 1-week retention measures, students retained 0.43 words per minute of instruction using TD and 0.46 words per minute of instruction using IR. Although these differences are not significant, an average of only 1.40 out of the six Esperanto words were retained 1 week post intervention using TD, and an average of 4.04 out of the six Esperanto words were retained 1 week post intervention using IR. Thus, while the number of words retained per minute of instruction was commensurate for TD and IR, overall, students were able to maintain more of the words taught using IR.

Burns and Sterling-Turner (2010) further demonstrated the usefulness of IR, specifically touting the importance of evaluating retention in addition to initial learning when calculating efficiency. Although previous research has supported the effectiveness of IR, prior outcomes regarding the efficiency of IR may have been underestimated since long term maintenance data were not considered in efficiency calculations. Thus, although TD produces favorable initial learning, IR may be more conducive to long-term maintenance, even though intervention sessions take more time. Since the effectiveness of an intervention is partially determined by its durability (Kratochwill & Stoiber, 2002), future research should consider maintenance measures, in addition to initial learning,

when determining the effectiveness and efficiency of drill and practice procedures (Burns & Sterling Turner, 2010; Skinner, 2010).

Improvements to the IR procedure. Many studies have demonstrated the effectiveness of IR (Burns & Boice, 2009; Burns & Kimosh, 2005; Burns & Sterling-Turner, 2010; Burns et al., 2004; MacQuarrie et al., 2002; Nist & Joseph, 2008; Schmidgall & Joseph, 2007). Rather than comparing IR to other drill procedures, researchers have started to compare the traditional IR procedure to adapted variations of IR in order to improve the procedure. In a recent study by Petersen-Brown and Burns (2011), a traditional IR procedure was compared to an IR procedure including a vocabulary component. Sixty-one students in second and third grade were taught a list of seven unknown words from the Fry Instant Word List using IR and again using IR with a vocabulary component. The IR condition was implemented as described previously by simply showing the words to the students using IR, asking them to repeat the word, and asking them to use the word in a sentence. IR with the vocabulary component was implemented the same way except students were also taught the definition of the word and were asked to say the word and definition each time it was presented. One week after instruction, students retained 93.2% of the words taught during the vocabulary condition and they retained 84.1% of the words taught during the traditional IR procedure. Additionally, students generalized 93.7% of the words in the vocabulary condition and 82.1% of the words during the traditional IR condition. Thus, although the traditional IR procedure was an effective method of teaching sight words to these students, the procedure was enhanced when a vocabulary component was added.

Kupzyk, Daly, and Anderson (2011) compared two variations of IR to evaluate the effect of manipulating some of the traditional features of IR. Using four first-grade students as participants, a counterbalanced A-B-A-B design was used to compare the two conditions. Students were taught sight words from the Dolch sight word list. The traditional IR procedure was implemented as described previously. During the modified IR procedure, called strategic incremental rehearsal (SIR) by the authors, experimenters used systematic prompting methods and manipulated when and how many times words were rehearsed based on student responses, rather than adhering to the traditional IR sequence outlined by Tucker (1989). Results suggested that although both methods resulted in high numbers of words read correctly, students correctly read more words during the SIR procedure than during the IR procedure. Maintenance also remained slightly higher during the SIR procedure. Thus, this study demonstrates another example of how the positive effects of IR were increased by slightly manipulating the procedure.

OTR within the IR Procedure

Varying OTR to Each Unknown Word within the IR Procedure

Once an unknown word cycles through the IR sequence, it is typically folded-in to the practice deck of flashcards as a known word. Thus, although the word is technically considered known at this point, students are still exposed to it as a known word in the IR sequence, thereby increasing the OTR to that word. However, once the final unknown word is introduced and incrementally rehearsed, the number of OTR to each unknown word is not equivalent. For example, if ten unknown words were taught to a student using an IR procedure containing 90% known words and 10% unknown words, the first unknown word would be practiced significantly more times than the tenth unknown

word. This is because the first unknown word is rehearsed as a known word while subsequent unknown words are introduced. If the session ended after the tenth word was incrementally rehearsed, the number of OTR to each unknown word in this example would be: U1=54, U2=53, U3=51, U4=48, U5=44, U6=39, U7=33, U8=26, U9=18, U10=9. The variations in OTR to each unknown word poses some important questions regarding the effectiveness and efficiency of IR as described in previous studies (Volpe, Mulé et al., 2011).

Nist and Joseph (2008) reportedly held OTR constant across conditions when comparing IR, interspersal, and traditional drill. However, because unknown words were folded-in as known words during the IR condition, additional OTR to these words were afforded. Additionally, the OTR to each unknown word in the IR condition varied by virtue of when it was introduced. For example, the first unknown word presented was practiced nine times as an unknown word. Then, after it was folded-in, students experienced additional OTR to this word while it occupied and moved through the first, second, third, fourth, fifth, and finally the sixth known word positions in the IR sequence. However, the final and sixth unknown word was practiced only nine times and was never placed in the known word position because the procedure ended at that point. The second, third, fourth, and fifth unknown words also had varying levels of OTR because they, too, were folded-in as known words. Thus, OTR was in fact, not held constant across conditions in the study by Nist and Joseph. Rather, the number of words taught across conditions was the constant variable (Volpe, Mulé et al., 2011). Moreover, these variations in OTR within the IR procedure were present in all IR studies reviewed previously in this literature review. Therefore, since precise measures of OTR are lacking

in the literature, inferences regarding the effectiveness and efficiency of IR, as compared to other drill and practice procedures, are somewhat equivocal.

Methodological Issues Effecting OTR in Previous IR Studies

A methodological oversight in the study by Joseph and Nist (2008) demonstrates another example of varying levels of OTR within IR studies. In this study, word retention was initially measured the day following each treatment condition. Words not retained on the next-day retention measures were reintroduced as unknown words into their respective treatment conditions (i.e., IR, TD, or interspersal) and practiced again. Thus, of the six unknown words taught during each session, it is possible that some of these words had been practiced the previous day, but not retained. Thus, since additional OTR to these words were provided by reintroducing them as unknown words for a second time, the authors' claim of holding OTR constant across conditions is confounded even further. While the article does not specify this level of detail regarding the reintroduced words, it is possible that the words not retained on next-day measures consisted of those that were presented toward the end of the IR cycle with the fewest OTR. This methodological oversight raises further concerns about Nist and Joseph's conclusions regarding the effectiveness and efficiency of IR and the other drill methods.

This limitation leads to another methodological issue in regard to when a word has been "mastered," or is considered "known." In the studies reviewed to this point, arbitrary rules for mastery, across a variety of drill and practice procedures, have been used. For example, in traditional drill and practice, words are typically considered "known" once they have been correctly responded to three times (Burns & Boice, 2009; Burns & Sterling Turner, 2010; MacQuarrie et al., 2002). For interspersal procedures, a

variety of mastery definitions have been established, depending on how the procedure was established (Burns & Boice, 2009; Cates et al., 2003; MacQuarrie et al., 2002). The IR procedure also contains an arbitrary mastery criteria suggesting that a word is "known" after it has cycled through the unknown word position of the IR sequence. However, several studies using IR have used fewer than nine known words (Bunn et al., 2005; Burns & Sterling Turner, 2010; Volpe, Mulé et al., 2011), creating even more variation in the mastery criteria for IR. Thus, a consistent mastery criterion for words taught during drill and practice procedures has not been established in the research.

The mastery criterion of IR is directly related to the OTR to each unknown word. Once an unknown word is folded-in, it is identified as "known," but the student is undoubtedly gaining something from the additional OTR. Researchers and practitioners alike must ask themselves this question: If learning is still occurring by providing additional OTR to a "known" stimulus, how can this stimulus be confidently labeled as "known?" While it is clear that words taught using IR become "known" at some point, it is unclear when this change actually occurs. Future research that more closely examines mastery definitions and criteria is needed, particularly for IR (Burns & Sterling-Turner, 2010). This research may help to determine an optimal level of OTR for use within IR, as well as for other drill and practice procedures.

Precisely Measuring OTR to Each Unknown Word in IR

While high levels of OTR may be a key causal mechanism of IR (Burns, 2007a; MacQuarrie et al., 2002; Szadokierski & Burns, 2008), the varying OTR to each unknown word within IR has generated only minimal research attention. Szadokierski and Burns (2008) were the first to consider how variations in OTR to unknown words

within the IR procedure impacts word recall. The authors conducted an exploratory analysis to "examine the relationship between the number of word presentations and the proportion of recall." (Szadokierski & Burns, 2008, p. 602). Twenty-seven fourth-grade participants were taught the pronunciation and English meaning of 36 unknown words from Esperanto international language across four conditions with varying OTR and drill ratio levels. Participants were taught nine words during each treatment condition. Known words for each condition consisted of sight words from the Dolch (1936) and Fry Instant Word (Fry & Kress, 2006) lists. Students met with the experimenter during five sessions, across 5 weeks, experiencing a different condition each week for the first 4 weeks. Word maintenance was measured 1 week later, prior to administering the new condition. Week 5 consisted only of a measure of recall for words presented during week 4.

To understand how the data for this analysis were obtained, a more complete description of each treatment condition is outlined next. Condition one consisted of high OTR and a high known to unknown ratio, with 90% known words and 10% unknown words. The nine unknown Esperanto words were incrementally rehearsed and folded-in as in previous IR studies. OTR to unknown words ranged from 9 to 53, with a mean of 35.67 response opportunities. Unknown words presented in condition two had low OTR with a high known to unknown ratio. To accomplish this, a 90% known to 10% unknown ratio was used, but the unknown words were not folded-in as is typical with IR. The same known words were used throughout the entire sequence. Thus, the OTR to each unknown word was exactly nine in this condition. For condition three, high OTR and a moderate known to unknown ratio, consisting of 50% known words and 50% unknown words, was used. Nine unknown words were split into groups of three, then, three known words were

added to each group of three making three decks of six words each (i.e., three known words and three unknown words). Unknown words were placed in the first, third, and fifth positions of the deck. A modified IR procedure was used, which included a folding-in component. Words were presented in this sequence: 1U, 1K, 2U, 2K, 3U, 3K. First, only 1U, 1K was presented. Next, 1U, 1K, 2U was presented. One word was added each time until all words in the sequence had been presented. Once all of the words were rehearsed in this sequence, the whole deck of six cards was presented five more times for a total of six sequences; then, the three unknown words were folded-in to replace the previous three known words. This process was repeated for the final three unknown words. OTR to unknown words ranged from 12 to 60, with a mean of 34. Consisting of low OTR and a moderate known to unknown ratio, condition four was similar to condition three, but the sets of six were repeated only two times, and unknown words were not folded-in. OTR to unknown words in condition four ranged from 4 to 10 with a mean of 7.33. Thus, across each condition, the number of OTR to each unknown word varied drastically, with 17 different levels of OTR to unknown words, ranging from 4 to 60.

Without regard to treatment condition, words with ten or fewer OTR were retained about 15% of the time, whereas words with 18 to 60 OTR were retained 50 to 60% of the time. This analysis, although admittedly exploratory in nature, yielded important findings. First, results demonstrated no significant differences in word retention for words presented 18 versus 60 times. Ebbinghaus (1885) described a phenomenon, now known as the learning curve, that indicates learning gradually evens out over time. The learning curve concept may have some application to the findings of

Szadokierski and Burns (2008). To explain their findings, the authors speculate a retention plateau effect at a certain OTR threshold. If this hypothesis is true, practitioners could adjust the number of known words in the IR sequence so that an optimal number of OTR to each unknown word is achieved. Knowing the optimal level of OTR would reduce over-practicing and improve the efficiency of IR. Future research in this area is needed.

A second major finding of this study is related to the OTR to unknown words in IR, and the order in which unknown words are introduced. In review, the analysis by Szadokierski and Burns (2008) found that words presented ten or fewer times, regardless of the treatment condition, were retained at significantly lower rates than words presented 18 times or more. In this study, as well as in previous studies examining IR, unknown words introduced towards the end of the IR session had significantly lower levels of OTR than words introduced towards the beginning of the session. This is due to the increased OTR afforded by the folding-in procedure. Although the analysis by Szadokierski and Burns included four different variations of IR, it is evident that the final unknown words to be introduced during the treatment session presented the lowest OTR and demonstrated the poorest maintenance. Thus, the folding-in component may not necessarily be the catalyst for success in IR; rather, it may serve as a method of providing sufficient OTR for optimal maintenance. Future research is needed to confirm this hypothesis.

Although session length was not precisely timed in the study by Szadokierski and Burns (2008), rough estimates of efficiency were discussed. Across the two high OTR conditions, an average of 4.52 words were retained with sessions lasting about 26 minutes. For the two low OTR conditions, an average of 1.21 words were retained, and

sessions lasted approximately 15 minutes. When efficiency was calculated, students learned approximately 0.17 words per minute of instruction in the high OTR conditions and approximately 0.08 words per minute of instruction in the low OTR conditions. These results further support the importance of providing high levels of OTR. However, it should be noted that these efficiency estimates were based on retention of words that were taught with significantly varying OTR to unknown words. Thus, while providing high levels of OTR, future researchers should take care to hold OTR to each unknown target word constant when calculating the efficiency of IR (Szadokierski & Burns, 2008).

In a recent study by Volpe, Mulé et al. (2011), the effectiveness and efficiency of IR and traditional drill and practice (TD) were compared using a multi-element alternating treatment design. This study, together with the study by Szadokierski and Burns (2008), are currently the only published studies to purposefully consider the varying OTR to unknown words within the IR procedure. Unlike Szadokierski and Burns, however, Volpe and colleagues took steps to ensure similar OTR to each unknown word when calculating efficiency. Four first-grade students in general education were taught high frequency words using IR and TD. Since previous studies failed to accurately hold OTR constant across conditions, and across unknown words within the IR sequence, a modified IR procedure was used to ensure consistent OTR. This was achieved by removing the folding-in component of IR, which is comparable to condition two in the study by Szadokierski and Burns.

Skinner (2008) has criticized the practice of holding OTR constant when comparing the efficiency of two or more interventions, and suggests that the inconsistent amount of time spent during instructional conditions may skew efficiency calculations.

He also points out the possibility of increased learning potential towards the beginning of a learning trial. Therefore, Volpe, Mulé et al. (2011) compared TD and IR, each drill procedure having a time held constant condition and an OTR held constant condition. Thus, students participated in four different conditions, with three new unknown words taught during each condition.

In the TD-OTR-held-constant-condition (TD-ORC), three unknown words were presented one after the other, and repeated five times. Thus, students had five OTR to each unknown word. In the TD-time-held-constant-condition (TD-TC), the three unknown words were repeatedly presented one after the other until 3 minutes had elapsed. Thus, in this condition, students were afforded as many OTR as possible within the 3 minute time period. The IR-OTR-held-constant-condition (IR-ORC) taught three unknown words using an IR sequence containing five known words and one unknown word. Contrary to typical IR protocol, however, once the word cycled through the IR sequence, it was not folded-in as a known word. Thus, no further practice was permitted for the unknown words because they were incrementally rehearsed with the same five known words throughout the entire procedure. Each word was responded to exactly five times as it cycled through the IR sequence. In the IR-time-held-constant-condition (IR-TC), similar procedures were used except the treatment was continued until 3 minutes had elapsed. Thus, in this condition, OTR to unknown words was higher because the procedure was repeated until the time ran out. Next-day retention and 1-week maintenance measures were used to compare the effectiveness and efficiency of the four treatment conditions.

When comparing the TD-TC and IR-TC conditions, significantly higher OTR were found for unknown words within the TD-TC condition because the entire 3 minutes was spent repeatedly practicing only the targeted unknown words. However, the OTR to unknown words in the IR-TC condition was much lower because known words, in addition to the targeted unknown words, were practiced. Within the allotted 3 minutes, the average OTR across the three words was 17.13 (*SD* = 2.90) for IR and 78.53 (*SD* = 29.93) for TD. Dividing these averages by three, since there were three words practiced, yields an approximate OTR to each unknown word of 5.71 for the IR-TC condition and 26.18 for the TD-TC condition.

Notwithstanding the significantly higher OTR within the TD-TC condition, next-day retention and 1-week maintenance was about the same for TD-TC and IR-TC conditions. Thus, when time was held constant, TD and IR were equally effective. Regarding efficiency, TD and IR were also equally efficient when calculated using both the next-day and 1-week maintenance measures when time was held constant.

When OTR were held constant, TD and IR were also found to be equally effective. However, the TD-ORC condition was significantly more efficient than the IR-ORC. This was true when using both the next-day and 1-week maintenance measures to calculate efficiency. These results were similar to previous studies that used next-day retention measures for efficiency calculations (MacQuarrie et al., 2002; Nist & Joseph, 2008). However, previous research typically found IR to be more effective than TD (Burns & Boice, 2009; Nist & Joseph, 2008), but this was not the case in the study by Volpe, Mulé et al. (2011). Since the authors removed the folding-in component of IR in

this study, students were not afforded the additional OTR to unknown words found in previous IR studies.

Volpe, Mulé et al. (2011) as well as Szadokierski and Burns (2008) addressed the issue of variable OTR to unknown words by removing the folding-in component of IR. While this method allowed the OTR to unknown words to remain constant, it significantly reduced the total number of OTR to unknown words, thus reducing the effectiveness of the intervention. As hypothesized by Szadokierski and Burns, the additional OTR afforded by the folding-in component of IR, rather than the folding-in component itself, may be a more likely explanation for the effectiveness of IR. Future research should examine other possible ways to ensure consistent OTR to each unknown word, without removing the folding-in component of IR. Although not empirically tested, a promising approach could involve measuring maintenance only for unknown words that have completely cycled through the IR sequence. This would ensure an equivalent OTR to each unknown word as it cycles through the unknown and known word positions of the IR sequence.

High Levels of OTR versus Other Possible Causal Mechanisms of IR

Psychologists have long accepted the idea that practice and learning demonstrate a positive relationship (Ebbinghaus, 1885). Although high levels of OTR is a widely accepted explanation for the effectiveness of IR, some authors suggest other features, such as distributed practice effects (Burns et al., 2004; MacQuarrie et al., 2002), increased reinforcement rates (Skinner, 2002), enhancement of students' sense of success (Neef et al., 1977, 1980), and response pace (Hawkins, Skinner, & Oliver, 2005; Robinson & Skinner, 2002) may be key components of the success of IR.

Nist and Joseph (2008) suggested that the incremental and folding-in component of IR, rather than high levels of OTR, may have been responsible for the success of IR in their study. However, this assertion is based on their claim that OTR were held constant between conditions. Since the OTR to each unknown word across conditions was in fact not held constant in the Nist and Joseph (2008) study, it is possible that OTR is a more salient factor of IR than the incrementally folding-in aspect of IR as they originally hypothesized. However, future research confirming this hypothesis is needed.

Volpe, Mulé et al. (2011) concluded that high levels of OTR are at least partially responsible for the effectiveness of IR in previous studies. Additionally, the authors hypothesize that there may be something related to the folding-in procedure, other than the high levels of OTR it facilitates, that makes IR effective. While this may be possible, their study did not provide OTR that even approximated the levels found to be effective by Szadokierski and Burns (2008). For example, words were practiced only five times in the IR-ORC condition and an average of 5.7 times in the IR-TC condition. Szadokierski and Burns indicated that words practiced 18 or more times were retained at a much higher rate than words practiced 10 or fewer times. Therefore, since the OTR to unknown words taught via IR in the study by Volpe, Mulé et al., was so low, it is understandable why there was little difference in word retention across the various conditions.

Optimal OTR Level

Although research widely supports the necessity of repetition, or OTR, in the learning process, few studies have evaluated the optimal number of times a word should be practiced to facilitate retention. Based on his studies of word recognition with first-grade students in the 1920's, Gates (1930) produced the first published work in this area.

His year long study yielded two major findings. First, the optimal number of times a word should be responded to is widely variable among individuals. Second, students with lower intelligence required more repetition than students with average or above average intelligence. Notwithstanding the significant methodological limitations of his study, Gates suggested an optimal OTR level of 35 for students with average intelligence. At the outer extreme, Gates suggested that students with IQ scores ranging from 60 to 69 need about 55 OTR to facilitate retention and students with IQ scores of 120 to 129 need about 20 OTR to facilitate retention. However, these estimates are based on word repetitions necessary within basal readers, not for drill and practice procedures.

In an attempt to further investigate the findings of Gates (1930), Hargis, Terhaar-Yonkers, Williams, and Reed (1988) investigated optimal OTR across two conditions: a flashcard drill method and word exposure within text. Fifteen students with learning disabilities and two students with mental retardation participated in the study. Several procedural and instructional concerns exist in this study, however. First, words were presented to the students only 3 times per day in both conditions. This presents a significant time delay between OTR. Second, although this study included a flashcard drill procedure, the target words were presented with "decoy" words, which were also unknown words. Thus, although students were exerting mental energy to also learn the "decoy" words, the effect of including these words was not considered or measured. Third, target words included both sight words and decodable words. Therefore, whether the word was learned by repetition or through decoding strategies is unknown. These significant methodological limitations point to questionable results and conclusions. Notwithstanding, the authors suggested an average optimal OTR level of 53 for words in

isolation (flashcard condition) and 46 for words presented within text. Even though participants with learning disabilities and intellectual disabilities were used in this study, correlations between OTR and intelligence were not provided.

With this very limited research base, the optimal level of OTR necessary within drill procedures is still unclear. As discussed previously, Szadokierski and Burns (2008) provided some indication of an optimal level of OTR within the IR procedure by pointing out little benefit of exceeding 18 OTR. However, the findings of Szadokierski and Burns are preliminary and call for several areas of future research. First, the authors suggest future research that measures maintenance, and even generalization, of skills further out than 1 week, possibly up to 30 days postintervention. It is unclear if the levels of OTR recommended by Szadokierski and Burns would remain stable if word maintenance were measured this far out. Also, these results were derived from a limited sample with minimal variations in age and levels of intelligence. As suggested by Gates (1930), the optimal level of OTR could be different for individuals with different levels of intelligence. Therefore, future research should also include a diverse age range and ability level in order to see if variations of optimal levels of OTR exist for different ages and levels of intelligence. Since educational time is valuable, practitioners should not waste time by providing too many OTR, nor should they risk the possibility of poor word maintenance by not providing enough OTR.

OTR and Intelligence

A variety of instructional procedures containing high levels of OTR have been successful with children with disabilities (Burns, 2007a; Chase & Symonds, 1992; Kamps et al., 1994; Skinner et al., 1989). For students with cognitive delays, interspersal

procedures (Neef et al., 1977; 1980) and IR (Burns, 2007a; Burns & Boice, 2009; Burns & Kimosh, 2005), both of which include high levels of OTR, have been shown to produce desirable reading outcomes. IR has also effectively increased reading skills (Burns, 2007b; Burns et al., 2004) and math-fact acquisition (Burns, 2005) for students with learning disabilities. However, despite the preponderance of research supporting the importance of OTR for students with disabilities, teachers often do not provide enough OTR to promote optimal learning (Sutherland & Wehby, 2001; Wilson, Majsterek, & Simmons, 1996).

The importance of high levels of OTR for children with intellectual disabilities was highlighted in a study by Burns (2007a). A 9-year-old male student, with moderate mental retardation, was taught sight words using IR in a high OTR condition and a moderate OTR condition. A drill ratio of 10% unknown words to 90% known words were used for both conditions. Results indicated that although both conditions were effective, the student retained more words in the high OTR condition than in the moderate OTR condition. In the moderate OTR condition, 40% to 60% of the words were retained, whereas 72% to 92% of the words in the high OTR condition were retained. Thus, although this study contained only one participant, the importance of high levels of OTR for children with intellectual disabilities is clearly illustrated.

Some researchers have suggested that IR may negate the effect of individual differences on word retention (Burns & Boice, 2009; MacQuarrie et al., 2002). For example, MacQuarrie et al. (2002) used the Peabody Picture Vocabulary Test to measure receptive vocabulary for each participant. Correlation coefficients were calculated between receptive vocabulary and word retention for each retention measure: 1 day, 2

days, 3 days, 7 days, and 30 days. In the IR condition, coefficients ranged from -.16 to .06. The other two conditions, traditional drill and drill sandwich (a type of interspersal procedure), yielded coefficients ranging from .08 to .32 and .16 to .25, respectively. These data suggest that verbal ability and word retention were more highly correlated in the traditional drill and interspersal conditions than they were in the IR condition. In fact, there was virtually no correlation between verbal ability and word retention in the IR condition. The authors hypothesized that when OTR are high, as in IR, the relationship between individual differences and intervention outcomes may have small or no correlations. If this hypothesis is true, school psychologists might reframe their role to include less measurement and identification of individual differences and more implementation of interventions that reduce individual differences (Daly & McCurdy, 2002). The analysis by MacQuarrie and colleagues was admittedly exploratory in nature; therefore, the authors suggested that future research evaluate the correlations between intelligence and word retention when using IR. Despite their high correlation, receptive vocabulary and intelligence are different constructs and the authors were unsure if they would demonstrate the same correlations with word retention.

Since intelligence has long been accepted as a factor related to learning, (Gates, 1930; Jensen, 1989), determining if IR truly does negate the effect of intelligence on word maintenance is a significant and worthy research endeavor. Burns and Boice (2009) replicated the study by MacQuarrie et al. (2002) to examine correlations between word retention and intelligence using students with below average intelligence. Twenty seventh- and eighth-grade students, with IQ scores ranging from 61 to 85 participated in this study. Each student was eligible for special education services either under the

category of learning disability or mental retardation. Words from the Esperanto international language were taught using IR, TD, and interspersal following procedures similar to MacQuarrie et al. Nine words were taught during each condition, and retention was measured 1 and 2 weeks later. The authors hypothesized that the IR condition would demonstrate the lowest correlation between IQ and retained words, and would also lead to the most words retained. As predicted, the correlation coefficients were lowest for the IR condition. The correlation coefficients at 1 and 2 weeks were .03 and .15 for the IR condition, .37 and .50 for the TD condition, and -.61 and -.45 for the interspersal condition. In addition, more words were retained on one and two week maintenance measures in the IR condition than in the other two conditions. The average number of words maintained at 1 and 2 weeks were 5.30 and 4.95 for the IR condition, 2.05 and 1.80 for the TD condition, and 1.60 and 1.15 for the interspersal condition. Participants in the study by MacQuarrie et al. retained slightly more words overall in each condition, but the correlation coefficients were very similar. Overall, these results suggest low to moderate correlations between word maintenance and intelligence in the TD condition, moderate to significant correlations between word maintenance and intelligence in the interspersal condition, and an insignificant relationship between word retention and intelligence in the IR condition. Thus, an effective intervention such as IR can successfully negate the effect of intelligence on word maintenance.

Notwithstanding the significance of these findings, the necessary level of OTR to unknown words is still unclear, particularly for students with delayed intelligence. For example, the study by Burns and Boice (2009)—using participants with below average intelligence—and the study by Burns and Sterling-Turner (2010)—using participants

with presumed average intelligence—each taught unknown Esperanto words using IR and TD. In the TD condition for both studies, words were considered mastered and were removed from the practice deck when the student correctly responded three times in a row. Students with delayed cognitive abilities (Burns & Boice, 2009) ended up responding to unknown words in the TD condition about the same amount of times as they did in the IR condition. However, in the study by Burns and Sterling-Turner, students with presumed average intelligence reached the mastery criterion in the TD condition much quicker, with significantly fewer OTR in the TD condition than were provided in the IR condition. Two relevant implications are apparent from these studies. First, in order to meet the TD mastery criteria, students with delayed intelligence required more OTR than children without disabilities. Thus, students with below average intelligence may require a higher level of OTR, at least during traditional drill and practice. A second, and related, important implication of the study by Burns and Boice has to do with the similar OTR to unknown words in the TD and IR conditions. Despite the similar OTR levels, significantly more words were retained in the IR condition than were retained in the TD condition. Thus, it is possible that something about IR, in addition to high levels of OTR, is responsible for word retention, particularly for children with below average intelligence.

Results from the study by Burns and Boice (2009) clearly indicated that IR was more effective than TD or interspersal for students with below average intelligence. These findings are in alignment with previous research comparing the effectiveness of IR and other drill methods using nondisabled participants. However, the efficiency of IR for students with below average intelligence seems to differ from previous studies using IR

with general education students (Joseph & Schisler, 2007; Nist & Joseph, 2008). On average, the IR condition took about 25 minutes, the TD condition took about 20 minutes, and the interspersal condition took about 15 minutes. Using the 2 week maintenance measure to calculate efficiency, students maintained 0.20 words per minute using IR, 0.10 words per minute using TD, and 0.08 words per minute using interspersal. These findings are important because they suggest that IR was more efficient than the other procedures for children with below average intelligence. However, efficiency estimates were calculated using 2-week maintenance data, whereas efficiency has historically been calculated using initial learning or next-day retention data (Cates et al., 2003; Joseph & Nist, 2006; Nist & Joseph, 2008), or more recently using 1-week maintenance data (Burns & Sterling-Turner, 2010). Although efficiency was not the main dependent variable, and sessions were not precisely timed, Burns and Boice were the first to calculate efficiency based on 2-week maintenance measures. Thus, it is unclear if the higher efficiency of IR in this study is because it has longer lasting effects than TD and interspersal, or if it is because IR is more efficient for children with below average intelligence. Future research should more precisely measure the efficiency of IR and other drill methods using students with varying degrees of intelligence based on 2-week, or even longer, maintenance measures. Research should also focus on determining if the optimal number of OTR within the IR procedure differs for students with average and below average intelligence. In light of the findings of Burns and Boice (2009) regarding the efficiency of IR for students with below average intelligence, as well as the low correlations between IQ and word retention, determining if Gates' (1930) optimal OTR

estimates for varying levels of intelligence are accurate is an important line of future research.

Summary

Several key areas of future research have emerged from this review of the literature, and are guideposts for the proposed study. First, researchers should consider the varying OTR to each unknown word due to the folding-in component of IR. Since the majority of previous studies comparing the effectiveness and/or efficiency of IR to other drill and practice procedures did not consider these variations, inferences regarding the effectiveness and efficiency of these methods are somewhat equivocal (Volpe, Mulé et al., 2011). The current study demonstrated procedures to ensure consistent OTR to each unknown word and more accurately measured the relationship between OTR and word retention. Second, determining the optimal number of OTR to unknown words within IR is a key area of future research because it is directly related to efficiency. For example, if too few OTR are provided, the student may experience poor word retention. However, if too many OTR are provided, words may be over-practiced, reducing the efficiency of the procedure. Thus, determining the lowest possible number of known words to rehearse unknown words with, and still ensure adequate retention, will help determine the optimal drill ratio and resulting optimal OTR within IR. In order to more accurately determine the effectiveness and efficiency of each condition, retention data from at least 1-week, or even up to 30 days postintervention, should be used as outcome measures. Third, research examining the relationship between optimal levels of OTR and intelligence, when OTR to unknown words is held constant, is needed. Previous research suggests that IR negates the effect of individual differences such as verbal ability and intelligence. However, when

using IR, it is still unknown if the optimal level of OTR is the same for individuals with average intelligence as it is for individuals with below average intelligence. Fourth, previous research has suggested the possibility of something other than high levels of OTR being responsible for word retention in IR. Some of these proposals include the actual folding-in component of IR, practice effects, and word placement within the IR sequence. Once the optimal level of OTR in IR is clearer, and a method for accurately accounting for the OTR to each unknown word is established, future researchers can compare IR to other drill methods, specifically focusing on factors other than OTR as possible contributory components.

Problem Statement

Previous studies identifying an optimal level of OTR across varying intelligence levels for drill and practice procedures is limited and contain significant methodological oversights. Furthermore, to date, no studies have examined the effectiveness and efficiency of IR with the folding-in component intact, while truly holding the OTR to each unknown word constant. Also, studies suggesting that IR negates the effect of individual differences on word retention are preliminary and in need of further investigation. Thus, the optimal drill ratio, and resulting OTR to each unknown word in IR, particularly across varying intelligence levels, is unclear.

Using a balance of participants with average and below average intelligence, the current study compared three IR instructional conditions, each containing a different number of known words and resulting OTR to each unknown word (8 Known words=44 OTR, 6 known words=27 OTR, and 4 known words=14 OTR), in order to determine which condition is the most effective and efficient. As with the recent studies by

Petersen-Brown and Burns (2011) and Kupzyk, Daly, and Anderson (2011), the purpose of this study is to compare variations of IR in order to determine how to improve the IR procedure. Of particular interest is the point at which word retention declines due to too few OTR, as well as the point at which word retention reaches a ceiling and additional OTR are unnecessary. Thus, this study is designed to determine the optimal OTR level within IR for students with below average and average intelligence. As the number of OTR necessary for adequate retention declines, the amount of required instructional time also declines. Thus, instructional efficiency is also a variable of interest in this study. The number of OTR required for optimal word retention and instructional efficiency will be evaluated for students with average and below average intelligence to see if the optimal OTR varies among these two populations. The independent variables for this study are the different levels of OTR found within each instructional condition and the intelligence level of the participants. The dependent variables are word retention based on 1- and 3-week retention measures, as well as instructional efficiency based on initial learning, 1-week, and 3-week measures.

Research Questions and Hypothesis Statements

When presented with three IR conditions, each containing a different number of known words and resulting OTR to each unknown word (high OTR condition – 8 known words=44 OTR, medium OTR condition – 6 known words=27 OTR, and low OTR condition – 4 known words=14 OTR):

1. *Which instructional condition leads to the highest number of words retained based on 3-week retention measures?* It was hypothesized that the high and medium OTR conditions would lead to the highest word retention. It was also

hypothesized that the mean word retention for these two conditions would not be significantly different, suggesting little benefit of providing more than 27 OTR.

2. W*hich instructional condition leads to the highest number of words retained per minute of instruction based on 3-week retention measures?* It was hypothesized that the medium OTR condition would lead to the highest number of words retained per minute of instruction based on 3-week retention measures.
3. *Which instructional condition is associated with the least forgetting across time when comparing 1-week retention measures to 3-week retention measures?* It was hypothesized that the least forgetting across time between the two retention measures would be found during the medium and high OTR conditions.
4. *Which instructional condition is associated with the least forgetting across time per minute of instruction when comparing 1-week retention measures to 3-week retention measures?* It was hypothesized that the medium OTR condition would be associated with the least forgetting across time per minute of instruction when comparing 1- and 3-week measures.
5. *Do participants with average IQ scores retain significantly more words (based on 3-week retention measures) than participants with below average IQ scores, as measured at each instructional condition?* It was hypothesized that the participants in the average IQ group would retain significantly more words than the participants in the below average IQ group during the low and medium OTR condition. However, it was hypothesized that there would not be a significant difference in word retention for the two groups during the high OTR condition.
6. *Do participants with average IQ scores retain significantly more words per*

minute of instruction (based on 3-week retention measures) than participants with below average IQ scores, as measured at each instructional condition? It was hypothesized that the average IQ group would retain significantly more words per minute of instruction than the average IQ group across all instructional conditions.

7. *Do participants with average IQ scores recall significantly more words than participants with below average IQ scores across the three measurement times (initial learning, 1-week retention measures, and 3-week retention measures)?* Participants in each group will initially learn six words. It was hypothesized that the participants in the average IQ group would retain significantly more words than the participants in the low IQ group (collapsed across condition).
8. W*hich instructional condition is most preferred by students?* It was hypothesized that the participants would prefer the low OTR condition because it took the least amount of time.

CHAPTER 2

METHODOLOGY

Design

This study used a 2X3X3 mixed factorial design. The between-groups factor consisted of students in special education with below average intelligence and students in general education with average intelligence. The within-subjects factors was the three instructional conditions and the number of words learned, retained at 1 week, and retained at 3 weeks in each of these conditions, as measured on three occasions (i.e., initial learning, 1-week retention measures, and 3-week retention measures).

An additional analysis was conducted using the same between-group factors and the same three instructional conditions. However, in this analysis, the number of words learned and retained per minute of instruction in each of the three conditions, as measured on three occasions (i.e., initial learning, 1-week retention measures, and 3-week retention measures) was the second within-subject factor.

Setting

The study took place in a public school located in the suburbs of a large metropolis in the southwestern United States. This school serves approximately 900 children in grades kindergarten through eight. During the 2010-2011 school year, approximately 58% of student body was enrolled in the free and reduced lunch program.

Five self-contained special education classrooms for students with intellectual disabilities are also located on the school campus. Students in need of this level of special education support are transported to the school from seven other schools throughout the district.

Each study session, occurred in a quiet and well-lit office on the school campus. The student and experimenter sat at a table next to each other, with a laptop computer directly in front of the student. The experimenter was positioned next to the student, within reach of the laptop computer.

Participants

G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) was used to calculate effect size by conducting an a priori power analysis for a split plot repeated measures ANOVA. Parameter estimates were taken from previous studies using IR. A violation of the sphericity assumption was assumed in the power analysis with power set to .95. Most previous studies comparing IR to other drill methods yielded effect sizes of around 1.0 or higher. Using a very conservative effect size of .20, there was a 95% chance of detecting an effect using 20 total participants. To be conservative, 24 total participants were recruited; 12 in general education classes with average intelligence, and 12 in special education classes with below average intelligence.

In order to select students for the low IQ group, a review of special education records was conducted to identify a pool of 20 students in sixth, seventh, and eighth grade with previous full scale IQ standard scores ranging from 55 to 77.5. Prior to reviewing special education records, parents of students who receive special education services, and who were suspected of having low intelligence, were asked to sign a consent form allowing the researcher to review their child's special education records

(see Appendix A). Once permission was received, and files were reviewed, those students whose previous IQ scores were between 55 and 77.5 were considered potential participants. Students in sixth through eighth grade were used as participants because students in the primary elementary grades who would have met the below average intelligence standard score inclusion criteria are typically still learning the names and sounds of letters, and would not have been able to read the target words or the known words.

An initial pool of 20 sixth, seventh, and eighth grade students in general education classes was established based on teacher nomination. Each teacher was given a letter explaining the purpose of the study and was asked to nominate three potential participants (see Appendix B). This letter specifically asked teachers to only recommend students with average academic skills and no suspected learning or cognitive delays.

Parents of the potential participants in both the special education and general education groups were contacted by the primary researcher. The purpose of the study, inclusion criteria, study procedures, and potential risks and benefits were explained to the parents. A parent permission letter (see Appendix C) explaining the background and purpose of the study, inclusion criteria, study procedures, potential risks and benefits, and contact information for the researcher and Institutional Review Board was provided to the parents. Once written parental permission was obtained, the study was verbally explained to the students and they were asked to sign a child assent form (see Appendix D).

Prior to being selected as a final participant, the Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman, 2004) was administered to each potential participant. The KBIT-2 is an individually administered brief measure of

intelligence, measuring both verbal and nonverbal abilities. It has satisfactory reliability and validity and correlates well with the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV; Wechsler, 2003) and the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). Since the KBIT-2 takes only about 15 to 30 minutes to administer, it is a useful measure of intelligence for screening purposes (Sattler, 2008). The KBIT-2 was used in order to standardize the measure of intelligence across participants, to ensure a standardized test administration protocol for all participants, and to ensure that full scale IQ standard scores fell within the participant selection criterion ranges. To be included in the study, the students in the special education group must have obtained a full scale standard score on the KBIT-2 between 55 and 77.5. Students in the general education group were required to obtain a full scale standard score on the KBIT-2 between 92.5 and 115. The first 12 special education students and the first 12 general education students with signed parent permission and child assent forms, and who met the full scale IQ standard score inclusion criteria, were included in the study.

During the third session of the intervention, one of the students in the special education group refused to participate. The student became very silly and responded with unrelated answers such as, “I am a flying clown.” When asked to settle down and participate, he responded by saying, “I don’t want to” and put his head down on the table. When the experimenter encouraged the student to participate, he again said, “I don’t want to,” keeping his head on the table. He was then dismissed to class. The following day, the experimenter again encouraged the student to participate and he declined. Therefore, 23 students actually completed the study procedures.

Of the 23 participants, 11 met state special education eligibility criteria for specific learning disability (n = 4) or mild intellectual disability (n = 7) with intelligence standard scores ranging from 61 to 77 (M = 68.18, SD = 6.82), as measured by the KBIT-2. These participants were receiving special education services either in a self-contained special education classroom or a resource classroom. The average KBIT-2 standard score for students in the self-contained special education classroom was 64.4 (SD = 4.68). Each of these participants was classified as having an intellectual disability. The average KBIT-2 standard score for students in the resource classroom was 71.5 (SD = 7.37). Each of these participants was classified as having a learning disability. Scores on the KBIT-2 for all of the participants in the low IQ group were commensurate with their previous IQ scores. The other 12 participants were in general education classes and were not eligible for special education services. These students had intelligence standard scores ranging from 93 to 109 (M = 101.00, SD = 6.63), as measured by the KBIT-2.

Among the 11 participants receiving special education services, 54.54% (n = 6) were male, and 45.46% (n = 5) were female. At the close of the study, the age of the participants ranged from 11 years, 7 months to 14 years, 11 months with a mean age of 13 years, 5 months and a median age of 13 years, 11 months. The participants included 36.36% (n = 4) sixth graders, 18.18% (n = 2) seventh graders, and 45.46% (n = 5) eighth graders. Furthermore, 45.46% (n = 5) were identified as Caucasian, 27.27% (n = 3) were identified as African American, and 27.27% (n = 3) were identified as Hispanic.

Among the 12 of participants in general education classes, 41.67% (n = 5) were male, and 58.33% (n = 7) were female. The age of the participants ranged from 11 years, 6 months to 14 years, 4 months with a mean age of 13 years, 0 months and a median age

of 13 years, 2 months. The participants included 33.33% (*n* = 4) sixth graders, 25.00% (*n* = 3) seventh graders, and 41.67% (*n* = 5) eighth graders. Furthermore, 50.00% (*n* = 6) were identified as Caucasian, 25.00% (*n* = 3) were identified as African American, and 25.00% (*n* = 3) were identified as Hispanic.

Instructional Materials

During each instructional condition, students were taught the pronunciation and English translation of six unknown words from the Esperanto international language. Esperanto was developed towards the end of the 19th century by Ludwig Zamenhof as a proposed international communication medium (Richardson, 1988). Although the language is still used by some, its popularity did not become widespread and Esperanto is rarely used in the United States (Janton, 1993). Esperanto and English words share similar word configurations due to their European roots. A few examples of Esperanto words are cervo (deer), stelo (star), and pilko (ball). As with previous studies by Burns and colleagues (Burns & Boice, 2009; Burns & Sterling Turner, 2010; MacQuarrie et al., 2002, Szadokierski & Burns, 2008), Esperanto words were used as unknown words since the likelihood of previous exposure to these words is extremely low. Also, Esperanto words were used to help control for any possible learning opportunities between the instructional session and measures of recall, which was a weakness identified in previous studies (Semb & Ellis, 1994). In order to retain consistency across words and conditions, each of the targeted Esperanto words was a noun and contained exactly five letters and two syllables. Rather than using traditional Esperanto phonology, Esperanto words were taught using conventional English pronunciation rules. Although using English words as unknown words may have led to increased external validity, the tight experimental

control afforded by using a standardized list of Esperanto words more strongly linked the dependent and independent variables, thus leading to increased internal validity (Burns & Sterling-Turner, 2010).

Of the 95 commonly used English nouns identified by Dolch (1936), 35 contain exactly four letters and one syllable. These words were included in a pool of possible known words to be used in this study. Prior to being included in the study as known words, students were required to correctly identify words within 3 seconds, on two consecutive presentations. As a result of this screening process, a list of 18 known Dolch words were selected and used in the study. The Dolch and Esperanto words did not have the same English meaning.

Previous research examining IR suggests that fifth-grade students can effectively learn approximately 6.6 new words during one session (Burns, 2004a), before they make three errors and reach their acquisition rate. While this may be true for students in general education classes with presumed average intelligence, the optimal acquisition rate for students with below average intelligence has not been examined. Although acquisition rates were not mentioned in their study, Burns and Boice (2009) taught nine words during each instructional session, using IR, to students with learning and intellectual disabilities. In another study using IR, Burns (2007a) taught a 9-year-old student with a moderate intellectual disability five words per session. However, this study did not address acquisition rates either.

To ensure that the current study did not teach too many target words per instructional session, a preliminary investigation of acquisition rate was conducted with a fifth-grade student with below average intelligence. Six target words were taught using

an IR condition with eight known words. This student did not make more than three errors per word during this preliminary investigation. Therefore, it was determined that six unknown Esperanto words would be taught during each instructional condition throughout the actual study. Eighteen Esperanto words were randomly assigned to one of three sets and can be found in Appendix E with their English translation. Appendix F lists the 35 potential known words from the Dolch list, as well as the 18 Dolch words actually used in the study.

Using Microsoft Office PowerPoint 2010, a laptop computer with a diagonal screen size of 14 inches was used to display words during each instructional session, and retention measure. Words were displayed in the center of the screen using 120 point Century Gothic font. This was the only visual stimuli presented on the screen. The instructional conditions and corresponding retention measures displayed a pale blue (High OTR Condition), yellow (Medium OTR Condition), or red (Low OTR Condition) background color so students were able to identify treatment preference by color at the end of the study. The laptop computer was positioned on a table directly in front of the student, with the screen positioned approximately 2 feet from their eyes. The researcher sat next to the student, within reach of the laptop computer. This study differs from the majority of previous studies using IR because words were presented on a computer screen rather than on index cards. A stopwatch was used to precisely measure the duration of each instructional session.

A data collection sheet was used during each instructional condition and retention measure. A sample data collection sheet is found in Appendix G. The spelling and English translation of the Esperanto words was displayed on this sheet to aid the

researcher during the sessions. The number of times a student incorrectly responded to target words during the instructional procedure was recorded on the data collection sheet. Instructional duration, 1-week word retention, 3-week word retention, and efficiency calculations were also recorded on this form. A script containing a list of procedural steps was used by the experimenters to ensure treatment fidelity and consistency across participants and instructional conditions (see Appendix H).

Procedure

Training of Experimenters

Each KBIT-2 assessment session, instructional session, and retention measurement session was facilitated by the primary researcher or one-of-two school psychology graduate students. The primary researcher provided a 1-hour training session to the graduate student experimenters outlining the purpose of the study and specific instructions for implementing each procedure. Each experimenter demonstrated 100% adherence to the procedures outlined in the script and procedural steps form (see Appendix H) during two mock instructional and retention sessions which were held prior to initiating the study. Although the experimenters had no previous experience implementing IR prior to this study, the use of standardized PowerPoint presentations reduced the need for extensive procedural training, as well as the potential for procedural error.

Session Content and Sequencing

Participants met one-on-one with an experimenter during six sessions, with each session spaced approximately 1 week apart. Sessions one through three consisted of one

of the three instructional conditions, based on random assignment. Sessions two through four consisted of retention measures of the words taught during the previous week's instructional condition. During sessions two and three, the retention measures occurred immediately before the new instructional condition started. Sessions four through six consisted of retention measures of the words taught during the instructional condition three weeks earlier. Thus, word retention was measured approximately 1 and 3 weeks after instruction. Due to student absences and field trips, the sessions did not always occur exactly 1 week apart. Across all participants and conditions, the 1-week retention measures occurred an average of 7.14 (*SD* = 1.10) days after the instructional session and the 3-week retention measures occurred an average of 21.07 (*SD* = 0.40) days after the instructional session. The sequence and contents of each session is clearly represented in Appendix I.

Random Assignment

The 18 unknown Esperanto words were randomly assigned to one of three sets, with each set containing six words. These sets were counterbalanced and randomly assigned across the instructional conditions. The order in which participants experienced the three instructional conditions was also randomly assigned and counterbalanced. Thus, although each participant experienced all three instructional conditions, the set of six Esperanto words taught during the instructional conditions varied. A total of nine instructional conditions and Esperanto set combinations were used and are outlined in Appendix J.

Instructional Conditions

Three IR conditions containing a different number of known words and resulting OTR to each unknown Esperanto word were compared. IR typically uses nine known words (Tucker, 1989). However, since no effectual differences were previously found using eight or nine known words (Burns, 2004b), this study used eight known words in the instructional condition with the most known words. The instructional conditions contained the following number of known words and resulting OTR to each unknown Esperanto word:

1. High OTR Condition – 8 known words and 44 OTR to each unknown Esperanto word
2. Medium OTR Condition – 6 known words and 27 OTR to each unknown Esperanto word
3. Low OTR Condition – 4 known words and 14 OTR to each unknown Esperanto word

The instructional conditions followed the sequence of words as outlined below.

1. High OTR Condition - U1, K1, U1, K1, K2, U1, K1, K2, K3, U1, K1, K2, K3, K4, U1, K1, K2, K3, K4, K5, U1, K1, K2, K3, K4, K5, K6, U1, K1, K2, K3, K4, K5, K6, K7, U1, K1, K2, K3, K4, K5, K6, K7, K8
2. Medium OTR Condition - U1, K1, U1, K1, K2, U1, K1, K2, K3, U1, K1, K2, K3, K4, U1, K1, K2, K3, K4, K5, U1, K1, K2, K3, K4, K5, K6
3. Low OTR Condition - , K1, U1, K1, K2, U1, K1, K2, K3, U1, K1, K2, K3, K4

U1 represents the first unknown Esperanto word. K1 represents the first known word; K2 represents the second known word, and so on. After the entire sequence was rehearsed as specified above for each condition, the second unknown Esperanto word was introduced into the U1 position, the first unknown Esperanto word was folded-in and placed in the K1 position, the known word previously occupying the K1 position was moved to the K2 position, and so on. The final known word was removed from the sequence each time an Esperanto word was folded-in. During each condition, the six Esperanto words were introduced in this manner until the final word was introduced and rehearsed. The IR procedure typically ends at this point in previous studies, which has caused variations in the OTR to each unknown word (Volpe, Mulé et al., 2011); thus, the current study included a procedure to ensure equivalent OTR to the six Esperanto words, without removing the folding-in component of IR. After the sixth unknown Esperanto word was introduced, the previously removed known words were re-introduced as if they were unknown words. This occurred until each of the Esperanto words cycled completely through the IR sequence. Thus, an equal number of OTR to each unknown Esperanto word was established within the respective instructional conditions. Re-introducing known words in this manner occurred during each condition. Therefore, the only difference between the three instructional conditions was the number of known words and resulting OTR to each unknown word.

Instructional Condition Procedures

The correct pronunciation and English translation of six Esperanto words was taught to the participants during each instructional condition. Words were taught using a PowerPoint presentation, displayed on a laptop computer positioned directly in front of

the participant. Since the complete sequence of unknown and known words for each condition were previously programmed within the PowerPoint presentations, implementation was simplified and standardized. Experimenters followed the procedures outlined on the script and procedural steps form during all instructional sessions (see Appendix H).

As each unknown Esperanto word was initially presented, and appeared for the first time on the PowerPoint presentation, the experimenter stated the word pronunciation and provided the English translation. For example, when introducing the Esperanto word "balen," the experimenter said, "This word is balen. Balen means whale." The student was then asked to restate the Esperanto word, its English translation, and use the word in a sentence. An example of an appropriate response is, "Balen means whale" followed by a sentence using the word whale appropriately such as "I saw a balen swimming in the ocean." Each time the Esperanto words were subsequently presented, even once they had been folded-in and presented as known words, students were asked to correctly pronounce and translate the word into English. The experimenter provided verbal reinforcement, such as "good job" or "that's right" for each correct response. If an incorrect response was given, or if the student failed to respond within 3 seconds, the experimenter immediately provided corrective feedback by stating the Esperanto word pronunciation and English translation (e.g., "This word is balen. Balen means whale.") After receiving corrective feedback, students were again asked to correctly state and define the word before moving on. Verbal reinforcement was provided for correct responses after corrective feedback had been provided. Each time a student made an error, the experimenter recorded a tally mark on the data collection sheet next to the

miscued word. When the known Dolch words were presented, students were asked to state the correct word pronunciation. Although only required a few times, corrective feedback was also provided for mispronounced known words. After each correct response, the experimenter immediately advanced to the next word in the IR sequence by pushing Enter or the right arrow key on the keyboard. A stopwatch was used to measure the duration of each session. It was started as soon as the first word was presented on the PowerPoint presentation, and was stopped once the sixth Esperanto word had completely cycled through the IR sequence and the PowerPoint presentation ended. The total session duration was recorded on the data collection sheet. Students were offered a small piece of candy, small toy, or sticker from a prize box after each instructional session.

Retention Measurement Procedures

Retention of Esperanto words was measured 1 and 3 weeks after each instructional condition. During sessions two and three, retention measures occurred prior to administering the new instructional condition. During the fourth session, the 1-week retention measure occurred prior to the 3-week retention measure. Using a PowerPoint presentation and laptop computer, the six Esperanto words were presented one at a time with the same background color used during the instructional condition. Esperanto words were considered retained if the student provided the correct pronunciation and English translation within 3 seconds. No error correction or feedback was provided during the retention measures. Data was recorded on the data collection sheet by either checking a box indicating the correct response, or by recording the incorrect response. Although these sessions were not timed, the retention measures took less than 1 minute for each

participant. Students were again offered a small reward from the prize box at the end of the retention measurement sessions during weeks four through six.

Efficiency Calculation Procedures

The efficiency of each instructional condition is expressed as a learning rate, or number of words learned and retained per minute of instruction. The duration of each instructional session was measured in seconds and was converted into minutes by dividing the number of seconds by 60 (e.g., 750 seconds equals 12.5 minutes, 750/60 = 12.5). This study yielded three estimates of efficiency by dividing the number of words initially learned (i.e., six), retained at 1 week, and retained at 3 weeks by the number of instructional minutes. For example, suppose condition one took 12.5 minutes and the student learned six words, retained five words at 1 week, and retained four words at 3 weeks. Efficiency would be reported by indicating that the student initially learned .48 words per minute of instruction (6/12.5 = .48), retained .40 words per minute of instruction at 1 week (5/12.5 = .40), and retained .32 words per minute of instruction at 3 weeks (4/12.5 = .32). These three efficiency calculations were recorded on the data collection sheet for each instructional condition.

Social Validity Measurement Procedures

After presenting the final 3-week retention measure during the sixth session, the students participated in a social validity interview with the experimenter. The experimenter asked each participant five questions and recorded their responses on the treatment acceptability form (see Appendix K). Specifically, students were asked:

1. Do you feel like the computer program helped you learn new words from the Esperanto language?
2. Did you enjoy learning the new words from the Esperanto language?
3. Did you like the session with the blue, yellow, or red computer screen the best?
4. Why did you like that color session the best?
5. Why didn't you like the other color sessions as much?

During this interview, if the student did not immediately respond to the question "Did you like the session with the blue, yellow, or red computer screen the best," or if they stated that they could not remember the difference between the conditions, the experimenter explained that the blue session had a high number of OTR and took the longest amount of time, the yellow session had a medium number of OTR and took less time than the blue session, and the red session had a low number of OTR and took the shortest amount of time.

Dependent and Independent Variables

The dependent variables of this study were: (1) the number of Esperanto words initially learned; (2) the number of Esperanto words retained on 1-week retention measures; (3) the number of Esperanto words retained on 3-week retention measures; (4) the number of Esperanto words initially learned per minute of instruction; (5) the number of Esperanto words retained 1 week after instruction per minute of instruction; and (6) the number of Esperanto words retained 3 weeks after instruction per minute of instruction.

The independent variables of this study were: (1) The different number of known words and resulting OTR to each unknown word across the three instructional conditions, and (2) the different intelligence levels of the participant groups.

Treatment Integrity

Treatment integrity was established using a school psychology graduate student functioning as an independent observer during 20% of the instructional sessions. A treatment integrity checklist (see Appendix L) was used to determine the percentage of procedural steps followed during each session. The independent observer placed a check mark next to each step followed. The treatment integrity checklist contained a total of 13 steps. The number of steps followed was divided by the total number of steps (i.e., 13) and multiplied by 100 to determine the percentage of steps followed.

Interscorer Agreement

Interscorer agreement was determined during 10% of the 1-week retention measurement sessions and 10% of the 3-week retention measurement sessions. A school psychology graduate student served as the independent observer. Using an interscorer agreement form (see Appendix M), the independent observer recorded responses by either checking a box indicating the correct response, or by recording the incorrect response. Esperanto words were considered retained if the student provided the correct pronunciation and English translation within 3 seconds. Retention ratings for each test item, as determined by both the independent observer and the experimenter, were then compared to determine the percentage of score agreements. Interscorer agreement was calculated by dividing the number of score agreements by the total number of words presented during the measure (i.e., six) and multiplying by 100. These calculations were recorded on the interscorer agreement form found in Appendix M.

The independent observer also performed efficiency calculations for the observed retention measurement sessions in order to ensure that efficiency calculations were

performed correctly and with fidelity. In order to calculate efficiency, the independent observer divided the number of words retained by the number of instructional minutes. The number of instructional minutes was transferred to the interscorer agreement form from the data collection sheet. The efficiency calculations of the independent observer were compared to the calculations of the experimenter to determine if the calculations matched. If they matched, a "yes" was recorded on the interscorer agreement form. If the calculations did not match, a "no" was recorded. In order to determine efficiency calculation agreement, the number of agreements was divided by the number of agreements plus disagreements and multiplied by 100.

Analysis

Data were analyzed using a repeated measures ANOVA. The within-subject factors for this analysis were the instructional conditions and the number of words initially learned, retained at 1 week, and retained at 3 weeks, as measured on three occasions (i.e., initial learning, 1-week retention measures, and 3-week retention measures). Intelligence level based on mean standard scores on the KBIT-2 served as the between-subject factor. Thus, this analysis produced three main effects: the main effect for instructional condition (each of the three conditions), the main effect for group (the below average IQ group and the average IQ group) and the main effect for time (initial learning, 1-week retention measures, and 3-week retention measures). Three two-way interactions were also produced: condition x group, condition x time, and group x time. A three-way interaction between condition, group, and time was also produced but was not significant.

An additional repeated measures ANOVA was conducted using the number of words initially learned, retained at 1 week, and retained at 3 weeks per minute of instruction as the dependent variable. The within-subject factors for this analysis were the instructional conditions and the efficiency measures, as measured on three occasions (i.e., the number of words initially learned per minute of instruction, the number of words retained per minute of instruction on 1-week retention measures, and the number of words retained per minute of instruction on 3-week retention measures). The between-subject factor was intelligence level. Thus, this analysis produced three main effects: the main effect for instructional condition (each of the three conditions), the main effect for group (the below average IQ group and the average IQ group) and the main effect for time (initial learning, 1-week retention measures, and 3-week retention measures). Three two-way interactions were also produced: condition x group, condition x time, and group x time. A three-way interaction between condition, group, and time was also produced but this interaction was not significant.

Pairwise comparisons between the conditions and time measures were conducted for both ANOVAs. As a measure of effect size, partial eta-squared was reported for both analyses to demonstrate the proportion of variance that the instructional conditions, intelligence level, recall across time, the interaction between the conditions and intelligence level, the interaction between conditions and time, and the interaction between time and intelligence level have on the number of words recalled (effectiveness) and the number of words recalled per minute of instruction (efficiency). The alpha level was set at .05. Additionally, Cohen's *d* (Cohen, 1988) was calculated to determine the magnitude of mean differences. Effect sizes were calculated by subtracting the two

means and dividing the difference by the pooled standard deviation. The pooled standard deviation was calculated by taking the square root of the average of the squared standard deviations. Cohen proposed that $d = 0.20$ represents a small effect, $d = 0.50$ represents a medium effect and $d = 0.80$ represents a large effect.

CHAPTER 3

RESULTS

This chapter begins with a brief overview of the study procedure and participants. Second, interscorer agreement and treatment integrity results are examined. Third, the assumptions regarding the repeated measures ANOVA are discussed. Finally, the results of each research question are presented.

Participants and Procedure

The purpose of this study was to compare the effectiveness (number of words recalled) and efficiency (number of words recalled per minute of instruction) of three variations of IR for students with average and below average intelligence.

Participants included 11 students receiving special education services in sixth through eighth grade with a mean IQ standard score of 68.18 (*SD* = 6.82) and 12 general education students in sixth through eighth grade with a mean IQ standard score of 101 (*SD* = 6.63), as measured by the KBIT-2. The entire participant sample included 11 males, 12 females, 8 sixth-graders, 5 seventh-graders, and 10 eighth-graders. Table 1 depicts the participant demographics. Students in special education were selected based on their IQ scores falling between 55 and 77.5 on previous IQ measures. Students in general education were selected based on teacher nomination. Before final selection, participants were required to obtain full scale IQ scores on the KBIT-2 that fell within a

Table 1

Participant Demographics

	Low IQ Group		Average IQ Group		All Participants	
	n	%	n	%	n	%
Male	6	54.54	5	41.67	11	47.83
Female	5	45.46	7	58.33	12	52.17
6th Grade	4	36.36	4	33.33	8	34.78
7th Grade	2	18.18	3	25.00	5	21.74
8th Grade	5	45.46	5	41.67	10	43.48
African American	3	27.27	3	25.00	6	26.09
Caucasian	5	45.46	6	50.00	11	47.83
Hispanic	3	27.27	3	25.00	6	26.09

predetermined range (55 to 77.5 for the special education group; 92.5 to 115 for the general education group). Students were taught the pronunciation and English translation of six Esperanto words during each of the three IR conditions. During the high, medium, and low OTR conditions, students were given 44, 27, and 14 opportunities to respond to each unknown word, respectively. Word retention was assessed 1 and 3 weeks later.

Interscorer Agreement

A school psychology graduate student recorded student responses on an interscorer agreement form (see Appendix M) during 10% of the 1-week retention

measurement sessions and 10% of the 3-week retention measurement sessions. This was done to ensure that the interventionist was properly recording and scoring the student responses. In order to be considered retained, students were required to correctly pronounce the Esperanto word and provide the English translation within 3 seconds. The ratings of the independent observer, and the interventionist were compared to determine the percentage of score agreements. Interscorer agreement was calculated by dividing the number of score agreements by the total number of words presented during the measure (i.e., six) and multiplying by 100. Results indicated 100% interscorer agreement for all observed retention sessions, suggesting excellent adherence to the scoring procedures.

During the same 1-week and 3-week recall probes, the school psychology graduate student also calculated and recorded the number of words retained per minute of instruction (efficiency) on the interscorer agreement form. Following the same procedure as the interventionist, the independent observer calculated efficiency by dividing the number of words retained by the number of instructional minutes (number of instructional minutes was transferred from the data collection sheet). This established a random check of efficiency calculations to ensure that the interventionist correctly followed the calculation procedures. Efficiency calculation agreement was determined by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. Results indicated 100% agreement for the efficiency calculations, suggesting excellent adherence to the efficiency calculation procedures.

Treatment Integrity

Using a treatment integrity checklist (see Appendix L), a school psychology graduate student conducted treatment integrity observations during 14 randomly selected

instructional sessions (20% of the instructional sessions). The observer placed a check mark next to each of the 13 steps that were correctly followed, as outlined on the treatment integrity checklist. This was done to ensure proper adherence to the treatment protocol by the interventionist. The total number of steps followed was divided by the total number of steps (i.e., 13) and multiplied by 100 to determine the percentage of steps followed. Across the 14 observations, 100% of the procedural steps were followed correctly, suggesting excellent treatment integrity.

Analysis of Variance Assumptions

Prior to evaluating the results, the assumptions of repeated measures ANOVA should be considered. First, data for the dependent variables should be normally distributed. Normality was assessed by investigating skewness and kurtosis. Although data were generally symmetrical and appropriately peaked, some skewness was detected during the low OTR condition. However, ANOVA is robust to minor violations of normality, providing there is equal sample size in each of the groups, and these findings are not concerning. The second assumption is homogeneity of variance. There were some deviations from normality in this sample. The *t* tests are sensitive to violation of this assumption. Therefore, corrected degrees of freedom are reported when the assumption of homogeneity of variance was violated. ANOVA is also robust to unequal variances; therefore, any small differences were not concerning. The third assumption is sphericity. For the ANOVA evaluating effectiveness, Mauchly's Test of Sphericity indicated that the assumption of sphericity was not met for time, therefore a Greenhouse-Geisser correction was used. For the ANOVA evaluating efficiency, Mauchly's Test of Sphericity indicated

that the assumption of sphericity was not met for any of the factors, therefore a Greenhouse-Geisser correction was used.

Research Questions

This study posed eight research questions addressing the effectiveness, efficiency, acceptability, and the effect of intelligence for three IR procedures. The results of each research question were hypothesized prior to implementing the study. Each research question and hypothesis is included next, followed by a narrative and graphical representation of the results.

Research Question One

Which instructional condition leads to the highest number of words retained based on 3-week retention measures? It was hypothesized that the high and medium OTR conditions would lead to the highest word retention. It was also hypothesized that the mean word retention for these two conditions would not be significantly different, suggesting little benefit of providing more than 27 OTR.

A within-subjects repeated measures ANOVA was performed using the number of words initially learned and retained at 1 and 3 weeks. This analysis indicated a significant main effect for condition $F(2, 42) = 10.46$, $p < .001$, $\eta_p^2 = .33$. Pairwise comparisons demonstrated significant differences between the high and low OTR conditions (Least Significant Difference, LSD, $p < .001$), as well as the medium and low OTR conditions (LSD, $p < .001$). A nonsignificant difference was found between the high and medium OTR conditions (LSD, $p = .925$).

Like the within-subjects analysis, a repeated measures ANOVA for the average IQ participant group yielded a significant main effect for condition $F(2, 22) = 7.16$, p .004, $\eta_p^2 = .39$. Pairwise comparisons demonstrated a significant difference between the high and low OTR conditions (LSD, $p = .007$), and the medium and low OTR conditions (LSD, $p = .011$). A nonsignificant difference was found between the high and medium OTR conditions (LSD, $p = .783$). The repeated measures ANOVA for the low IQ participant group also yielded a significant main effect for condition $F(2, 20) = 3.67$, p = .044, $\eta_p^2 = .30$. For this group, pairwise comparisons again demonstrated a significant difference between the high and low OTR conditions (LSD, $p = .02$), and the medium and low OTR conditions (LSD, $p = .04$). As with the average IQ group, a nonsignificant difference was found between the high and medium OTR conditions (LSD, $p = .866$).

When evaluating the data for the two participant groups and for the participants as a whole, the difference between the high and medium OTR conditions was not significant, suggesting essentially the same rate of recall for the two OTR levels. However, these analyses were collapsed across time, meaning that the number of words initially learned, retained at 1 week, and retained at 3 weeks were averaged. Therefore, paired-samples t tests were conducted to evaluate the difference between each of the conditions based on 3-week retention. Overall, students retained significantly more words during the high OTR condition than they did during the low OTR condition t (22) = 3.23, $p < .05$. Students also retained significantly more words during the medium OTR condition than they did during the low OTR condition t (22) = 3.35, $p < .05$. However, there were not significant differences found between the high and medium OTR conditions t (22) = .12, $p = .91$. This suggests that students retained a similar number of

words during these conditions. For the average IQ group, students retained significantly more words during the high OTR condition than they did during the low OTR condition t (11) = 2.98, $p < .05$. The average IQ group likewise retained significantly more words during the medium OTR condition than they did during the low OTR condition t (11) = 3.19, $p < .05$. However, there were not significant differences found between the high and medium OTR conditions t (11) = .27, $p = .79$, suggesting that the students in the average IQ group also retained a similar number of words during these two conditions. For the low IQ group, the paired-samples t test results were dissimilar to those for the average IQ group. Although the students in the low IQ group did retain more words during the high OTR condition than they did during the low OTR condition, there were not significant differences found between these two conditions, t (10) = 1.64, $p = .13$. Also, the difference between the medium OTR condition and the low OTR condition was not significant for this group t (10) = 1.49, $p = .17$. Similar to the average IQ group, there were not significant differences found between the high and medium OTR conditions t (10) = .20, $p = .85$.

In reviewing the retention averages for the two participant groups, the 3-week retention averages for the low IQ group were notably lower than the 1-week retention averages. However, the average 1 and 3 week retention for the average IQ group remained relatively consistent. Therefore, paired-samples t tests were also conducted to evaluate the difference between each of the conditions based on 1-week retention measures. For the participants as a whole, the students retained significantly more words during the high OTR condition than they did during the low OTR condition, t (22) = 4.66, $p < .001$. They also retained significantly more words during the medium OTR condition

than they did during the low OTR condition t (22) = 3.94, $p < .001$. The difference in word retention between the high and medium OTR conditions was not significant, t (22) = .10, $p = .92$. For the average IQ group, students retained significantly more words during the high OTR condition than they did during the low OTR condition t (11) = 3.49, $p < .05$. They also retained significantly more words during the medium OTR condition than they did during the low OTR condition t (11) = 2.73, $p < .05$. However, there were not significant differences found between the high and medium OTR conditions t (11) = .28, $p = .78$. In contrast to the paired-samples t tests using the 3-week retention data, when evaluating the 1-week retention data, the students in the low IQ group retained significantly more words in the high OTR condition than they did during the low OTR condition, t (10) = 3.07, $p < .05$. The low IQ group also retained significantly more words during the medium OTR condition than they did during the low OTR condition, t (10) = 2.83, $p < .05$. Word retention was not significantly different in the high and medium OTR conditions for the low IQ group, t (10) = .13, $p = .87$. Therefore, on the 1-week retention measures, the paired-samples t tests demonstrated similar results for the two participant groups.

Cohen's (1988) d effect sizes were calculated using the mean number of words retained during each condition in order to determine the magnitude of effect for each condition. Since the paired-samples t tests demonstrated different results when evaluating the 1-week retention measures, Cohen's d effect sizes were calculated using both the 1-week retention measures and the 3-week retention measures during each condition. Table 2 displays the mean number of words retained 3 weeks after each instructional condition.

Table 2

Number of Words Retained at 3 Weeks

	Low IQ Group		Average IQ Group		All Participants	
	Mean	SD	Mean	SD	Mean	SD
High OTR	1.27	1.19	3.33	1.92	2.35	1.90
Medium OTR	1.36	.81	3.17	1.90	2.30	1.72
Low OTR	.64	1.03	1.17	1.64	.91	1.38

Data for each participant group and the participants as a whole are shown. These data were used to calculate the effect sizes for 3-week retention.

When considering the 3-week retention data for the participants as a whole, a large effect size coefficient of $d = .88$ was found between the high OTR condition and the low OTR condition. A large effect size coefficient of $d = .89$ was also found between the medium OTR condition and the low OTR condition. However, a small effect size coefficient of $d = .03$ was found between the high OTR condition and the medium OTR condition.

When considering the 3-week retention data for the participants in the average IQ group, a large effect size coefficient of $d = 1.21$ was found between the high OTR condition and the low OTR condition. A large effect size coefficient of $d = 1.13$ was also found between the medium OTR condition and the low OTR condition. However, a small effect size coefficient of $d = .08$ was found between the high OTR condition and the medium OTR condition.

When considering the 3-week retention data for the participants in the low IQ group, a medium effect size coefficient of $d = .57$ was found between the high OTR condition and the low OTR condition. A medium to large effect size coefficient of $d = .78$ was found between the medium OTR condition and the low OTR condition. However, a small effect size coefficient of $d = .09$ was found between the high OTR condition and the medium OTR condition. Table 3 displays the effect size coefficients for each participant group and the participants as a whole using the 3-week retention means.

Table 4 displays the mean number of words retained 1-week after each instructional condition. Data for each participant group and the participants as a whole are shown. These data were used to calculate the effect sizes for the 1-week retention measures.

When considering the 1-week retention data for the participants as a whole, a large effect size coefficient of $d = 1.00$ was found between the high OTR condition and the low OTR condition. A large effect size coefficient of $d = 1.08$ was also found between the medium OTR condition and the low OTR condition. However, a small effect size coefficient of $d = .02$ was found between the high OTR condition and the medium OTR condition.

When considering the 1-week retention data for the participants in the average IQ group, a large effect size coefficient of $d = 1.14$ was found between the high OTR condition and the low OTR condition. A large effect size coefficient of $d = 1.06$ was also found between the medium OTR condition and the low OTR condition. However, a small effect size coefficient of $d = .10$ was found between the high OTR condition and the medium OTR condition.

Table 3

Cohen's d *Effect Size Coefficients – Based on 3-Week Retention Data*

	Low IQ	Average IQ	All Participants
High vs. Low OTR	$d = .57$	$d = 1.21$	$d = .88$
Medium vs. Low OTR	$d = .78$	$d = 1.13$	$d = .89$
High vs. Medium OTR	$d = .09$	$d = .08$	$d = .03$

Table 4

Number of Words Retained at 1 Week

	Low IQ Group		Average IQ Group		All Participants	
	Mean	SD	Mean	SD	Mean	SD
High OTR	2.27	1.95	3.67	1.78	3.00	1.95
Medium OTR	2.36	1.36	3.50	1.73	2.96	1.64
Low OTR	.82	1.17	1.58	1.88	1.22	1.59

When considering the 1-week retention data for the participants in the low IQ group, a large effect size coefficient of $d = .90$ was found between the high OTR condition and the low OTR condition. A large effect size coefficient of $d = .1.21$ was found between the medium OTR condition and the low OTR condition. However, a small effect size coefficient of $d = .05$ was found between the high OTR condition and the medium OTR condition. Table 5 displays the effect size coefficients for each participant group and the participants as a whole using the 1-week retention means.

Table 5

Cohen's d *Effect Size Coefficients – Based on 1-Week Retention Data*

	Low IQ	Average IQ	All Participants
High vs. Low OTR	.90	1.14	1.00
Medium vs. Low OTR	1.21	1.06	1.08
High vs. Medium OTR	.05	.10	.02

In reference to the research question, the high and medium OTR conditions both led to the highest word retention using 3-week retention measures for the participants as a whole. As hypothesized, the mean 3-week retention for these two conditions was not significantly different. Therefore, participants did not retain significantly more words during the condition with 44 OTR than they did during the condition with 27 OTR. These findings were also true for the average IQ group. However, for the participants in the low IQ group, the difference in 3-week retention was not significantly different across any of the conditions. On the other hand, when the 1-week retention averages were examined, the difference between the high and medium OTR conditions was not significant, but the differences between the high and low OTR condition, as well as the medium and low OTR conditions were significant.

Research Question Two

W*hich instructional condition leads to the highest number of words retained per minute of instruction based on 3-week retention measures?* It was hypothesized that the

medium OTR condition would lead to the highest number of words retained per minute of instruction based on 3-week retention measures.

An additional within-subjects repeated measures ANOVA was performed using the number of words initially learned and retained at 1 and 3 weeks per minute of instruction. This analysis indicated a significant main effect for condition $F(1.56, 32.85) = 17.42$, $p < .001$, $\eta_p^2 = .43$. A significant main effect for condition was also found for the low IQ group $F(2, 20) = 14.85$, $p < .001$, $\eta_p^2 = .60$ and the average IQ group $F(2, 22) = 8.53$, $p = .002$, $\eta_p^2 = .44$.

Pairwise comparisons for condition were not made since time was collapsed across condition. This means that the number of words initially learned and retained per minute of instruction would have been averaged across condition and then compared. As illustrated by Burns and Sterling-Turner (2010), instructional efficiency can greatly vary depending on if it is calculated using initial learning or retention data. Therefore, pairwise comparisons would not have accurately represented the efficiency of the three instructional conditions.

For the participants as a whole, students initially learned the most words per minute of instruction during the low OTR condition ($M = .81$, $SD = .22$), followed by the medium ($M = .47$, $SD = .14$) and high ($M = .32$, $SD = .09$) OTR conditions. However, when considering efficiency using the 3-week retention data, the medium OTR condition ($M = .20$, $SD = .18$) yielded the highest number of words retained per minute of instruction, followed by the low ($M = .13$, $SD = .21$) and high ($M = .13$, $SD = .12$) OTR conditions. When considering efficiency using the 1-week retention data, similar results were found. The medium OTR condition ($M = .24$, $SD = .16$) yielded the highest number

of words retained per minute of instruction, followed by the high (M = .16, SD = .12) and low (M = .17, SD = .24) OTR conditions.

Paired-samples t tests were conducted to evaluate the difference in 1- and 3-week retention per minute of instruction for the three conditions. The mean number of words retained at 3 weeks per minute of instruction during the high and medium OTR conditions was significantly different, t (22) = 2.45, p <.05. However, the mean number of words retained at 3 weeks per minute of instruction during the high and low OTR conditions was not significantly different, t (22) = .10, p = .94. This was also true for the medium and low OTR conditions, t (22) = 1.54, p = .14. When considering efficiency using the 1-week retention averages, the difference between the high and medium OTR conditions was significant, t (22) = 2.80, p <.05. However, as with 3-week retention per minute of instruction, the difference between the high and low OTR conditions was not significant, t (22) = .06, p = .95. Also, the mean number of words retained per minute of instruction during the medium and low OTR conditions was not significantly different, t (22) = 1.45, p = .16. Thus, although research question number one found the medium and high OTR conditions to be equally effective, the medium OTR condition was significantly more time efficient than the high OTR condition when considering both the 1- and 3-week retention data.

As hypothesized, the medium OTR condition led to the highest number of words retained at 3 weeks per minute of instruction. However, the difference between the medium and low OTR condition was not statistically significant. Therefore, these results should be interpreted in connection with the findings of research question number one. Students retained approximately the same number of words during the medium and high

OTR conditions, but they spent significantly less time learning these words during the medium OTR condition. In addition, students retained significantly fewer words during the low OTR condition than they did during the medium or high OTR conditions. Thus, the medium OTR condition was the most time efficient. When evaluated using the 1-week retention data, the same results were found. Therefore, since the medium OTR condition was one of the most effective procedures as well as one of the most time efficient procedures, it seems to be the most ideal procedure for use in educational practice.

Table 6 displays the mean number of words initially learned and retained per minute of instruction across the three conditions. The average number of instruction minutes for each condition is also included in this table.

Research Question Three

Which instructional condition is associated with the least forgetting across time when comparing 1-week retention measures to 3-week retention measures? It was hypothesized that the smallest amount of forgetting between the two retention measures would be found during the medium and high OTR conditions.

A within-subjects repeated measures ANOVA was performed using the number of words initially learned, retained at 1 week, and retained at 3 weeks. This analysis indicated a significant main effect for time $F(1.38, 29.04) = 258.94$, $p < .001$, $\eta_p^2 = .93$. Pairwise comparisons demonstrated significant differences between initial learning and 1-week retention measures (LSD, $p < .001$) as well as between initial learning and 3-week retention measures (LSD, $p < .001$). This significance level was expected since the likelihood of forgetting some words 1 and 3 weeks after initially acquiring them was

high. However, pairwise comparisons also demonstrated significant differences between 1-week retention measures and 3-week retention measures (LSD, $p < .001$), suggesting

Table 6

Number of Words Initially Learned and Retained Per Minute of Instruction

	High OTR	Med. OTR	Low OTR
	Mean (SD)	Mean (SD)	Mean (SD)
Low IQ Group			
Initial learning	.27 (.08)	.37 (.10)	.65 (.18)
1-Week Retention	.09 (.08)	.15 (.09)	.07 (.11)
3-Week Retention	.05 (.04)	.09 (.06)	.07 (.10)
Session Duration in Minutes	23.95 (6.09)	17.16 (4.82)	9.98 (3.02)
Average IQ Group			
Initial learning	.38 (.05)	.57 (.09)	.95 (.14)
1-Week Retention	.23 (.11)	.33 (.17)	.25 (.30)
3-Week Retention	.21 (.12)	.30 (.19)	.18 (.26)
Session Duration in Minutes	16.35 (2.68)	10.82 (1.62)	6.42 (.98)
All Participants			
Initial learning	.32 (.09)	.47 (.14)	.81 (.22)
1-Week Retention	.16 (.12)	.24 (.16)	.17 (.24)
3-Week Retention	.13 (.12)	.20 (.18)	.13 (.21)
Session Duration in Minutes	19.98 (5.96)	13.86 (4.73)	8.13 (2.82)

that a significant number of words were also forgotten between the 1- and 3-week probes. Although this does not specify during which condition the most significant forgetting across time occurred, or if there were differences between the two participant groups, it does indicate significant forgetting over time.

A significant two-way interaction between condition and time $F(4, 84) = 8.16$, $p < .001$, $\eta_p^2 = .28$ was found, suggesting that remembering words over time is related to the instructional conditions, or the number of OTR found within the conditions. Figure 1 depicts this interaction graphically.

Based on a visual inspection of these data, the mean number of words retained at 1 and 3 weeks during the high and medium OTR conditions was nearly identical. Also, the average number of words retained at 1 and 3 weeks during the high and medium OTR conditions was higher than the average number of words retained at 1 and 3 weeks during the low OTR condition. In order to determine if significant forgetting occurred between the 1- and 3-week probes, paired-samples *t* tests were conducted using the average number of words retained at 1 week compared to the average number of words retained at 3 weeks during each condition. Students forgot a significant number of words between the 1- and 3-week retention measures during the high OTR condition, $t(22) = 3.35$, $p < .05$, and the medium OTR condition $t(22) = 3.19$, $p < .05$. However, during the low OTR condition, the number of words retained at the two measurement times was not significantly different, $t(22) = 1.50$, $p = .15$.

A significant two-way interaction between group and time $F(1.38, 29.04) = 7.29$, $p = .006$, $\eta_p^2 = .26$ was found, suggesting that the rate of remembering across time for the two participant groups was not the same. Therefore, the difference between the mean

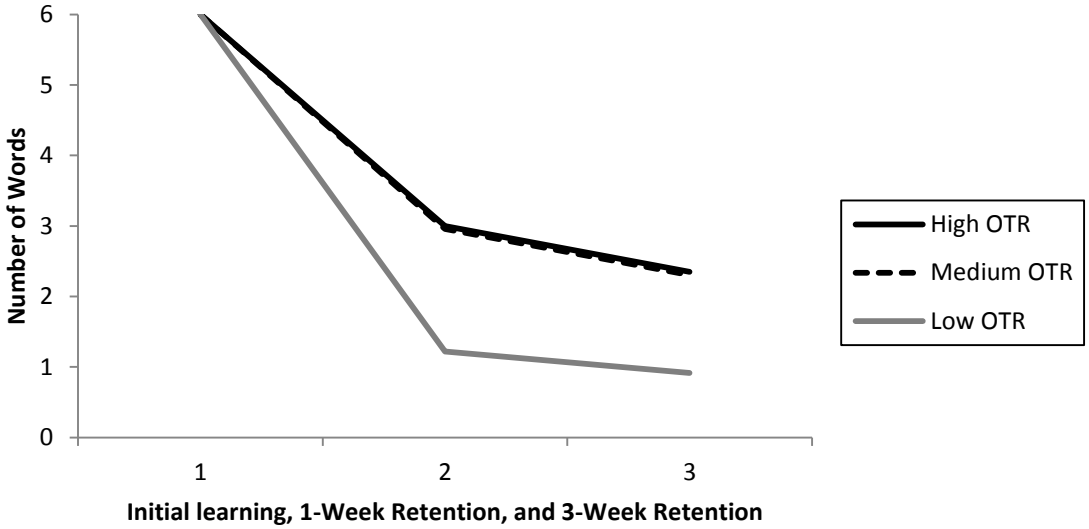

Figure 1. Two-way interaction between condition and time for all participants – Effectiveness.

number of words retained at 1 and 3 weeks was evaluated for the groups independently. Figure 2 depicts the interaction between group and time graphically.

For the average IQ group, a significant main effect for time $F(1.26, 13.84) = 73.44$, $p < .001$, $\eta_p^2 = .87$ was found. Pairwise comparisons demonstrated significant differences between initial learning and 1-week retention measures (LSD, $p < .001$) as well as between initial learning and 3-week retention measures (LSD, $p < .001$). Pairwise comparisons between 1- and 3-week retention measures were also significant (LSD, $p = .035$), suggesting that the students in the average IQ group forgot a significant number of words between the 1- and 3-week probes, collapsed across condition. A significant two-way interaction between condition and time $F(2.42, 26.62) = 6.01$, $p = .005$, $\eta_p^2 = .35$ was also found suggesting that differences in word recall depended on the instructional

condition. As with the participants as a whole, paired-samples *t* tests were also conducted for the average IQ group to determine if significant forgetting occurred between the 1-

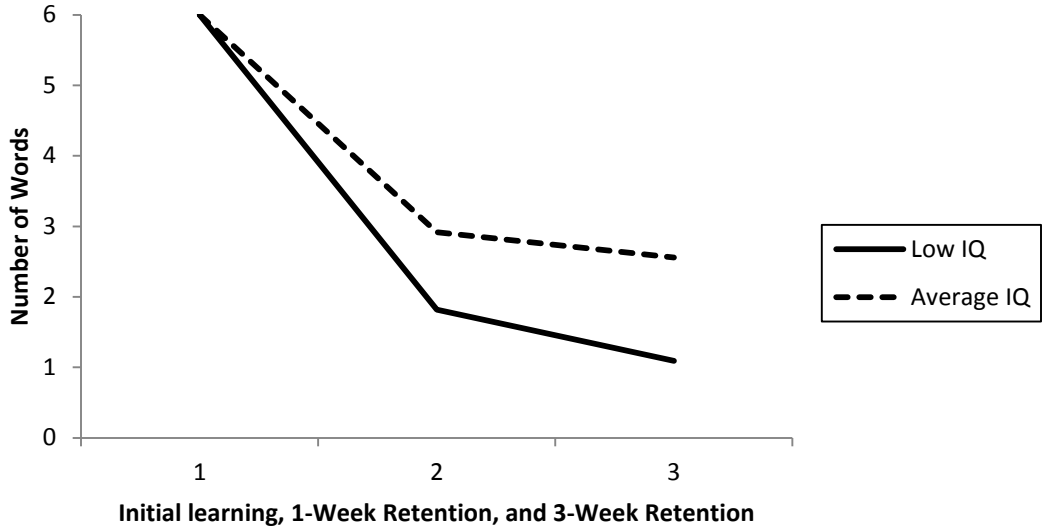

Figure 2. Two-way interaction between group and time – Effectiveness.

and 3-week probes. Students in the average IQ group forgot a significant number of words between the 1- and 3-week retention measures during the high OTR condition, t (11) = 2.35, $p < .05$. However, the number of words retained at 1 and 3 weeks was not significantly different during the medium t (11) = 1.48, $p = .17$ or low OTR conditions t (11) = 1.33, $p = .21$.

For the low IQ group, a significant main effect for time $F(1.2, 12.01) = 244.49$ $p < .001$, $\eta_p^2 = .96$ was found. Pairwise comparisons demonstrated significant differences between initial learning and 1-week retention measures (LSD, $p < .001$) as well as between initial learning and 3-week retention measures (LSD, $p < .001$). Pairwise comparisons between 1- and 3-week retention measures were also significant (LSD, $p = .005$), suggesting that the students in the low IQ group also forgot a significant number of words between the 1- and 3-week probes. A significant two-way interaction between

condition and time $F(4, 40) = 3.18$, $p < .023$, $\eta_p^2 = .24$ was found for the low IQ group as well suggesting that differences in word recall depended on the instructional condition for this participant group, too. Paired-samples *t* tests were also conducted for the low IQ group to determine the significance of forgetting across time during each condition. Students forgot a significant number of words between the 1-week retention and 3-week retention measures during the high OTR condition, $t(10) = 2.80$, $p < .05$ and the medium OTR condition $t(10) = 3.03$, $p < .05$. However, during the low OTR condition, the number of words retained at 1 and 3 weeks was not significantly different, $t(10) = .69$, $p = .51$.

Figure 3 depicts the two-way interaction between condition and time for the average IQ group and Figure 4 depicts this interaction for the low IQ group.

In order to determine which instructional condition was associated with the smallest difference in word recall across time, Cohen's (1988) *d* effect sizes were computed. The mean number of words retained on 1-week retention measures were compared to the mean number of words retained on 3-week retention measures. This was calculated for both participant groups and the participants as a whole for each condition. The objective was to determine which conditions demonstrated small differences when comparing 1- and 3-week measures, and to see if there were varying results among the two groups. Instructional conditions with small effect size coefficients represent OTR levels that facilitated better long term retention and small differences in word recall between the 1- and 3-week probes. Alternatively, conditions with large effect size coefficients represent OTR levels with large differences in word recall between the 1- and 3-week probes suggesting poor long-term retention. Table 7 displays the mean

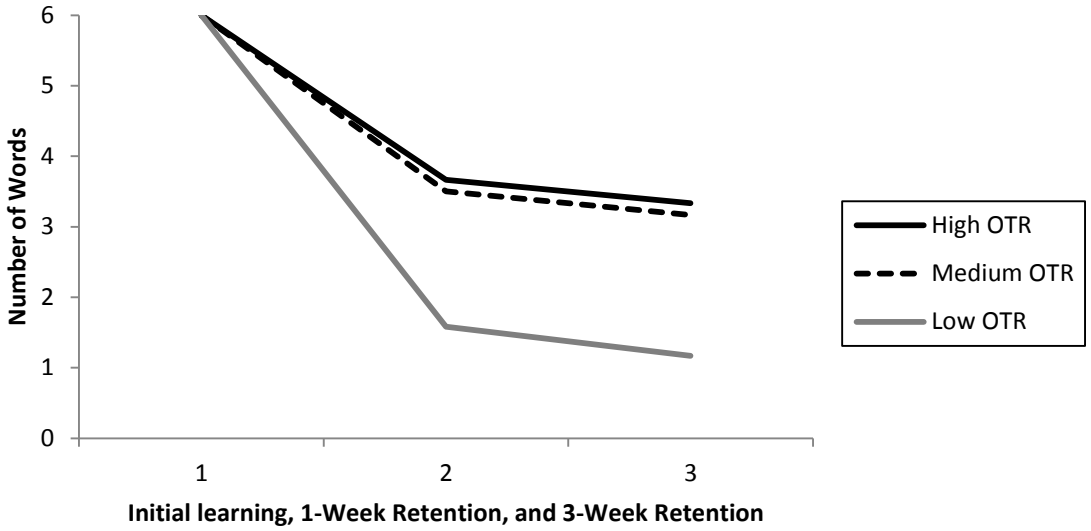

Figure 3. Two-way interaction between condition and time for the average IQ group – Effectiveness.

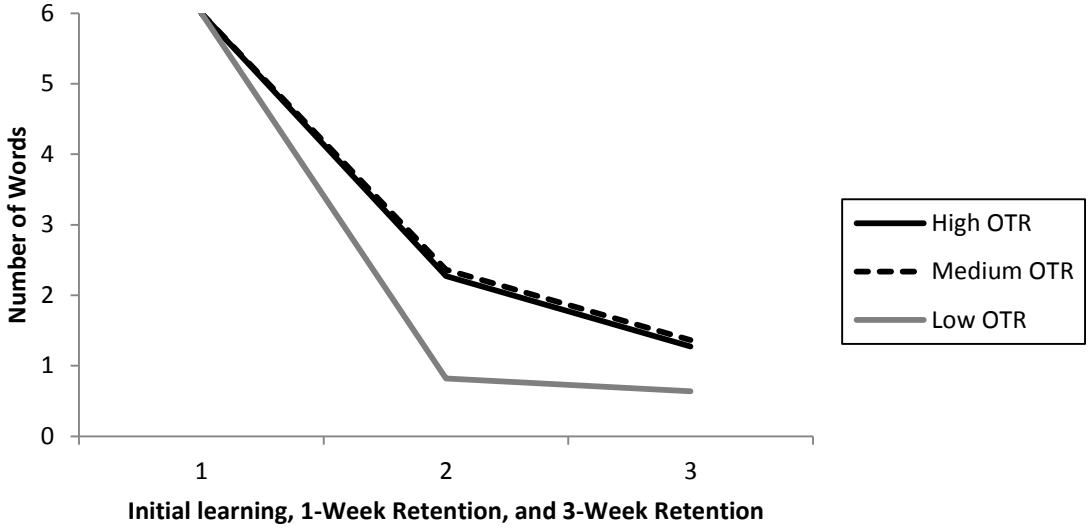

Figure 4. Two-way interaction between condition and time for the low IQ group – Effectiveness.

Table 7

Number of Words Retained at 1 and 3 Weeks

	Low IQ Group		Average IQ Group		All Participants	
	1 Week	3 Weeks	1 Week	3 Weeks	1 Week	3 Weeks
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
High OTR	2.27 (1.95)	1.27 (1.19)	3.67 (1.78)	3.33 (1.92)	3.00 (1.95)	2.35 (1.90)
Medium OTR	2.36 (1.36)	1.36 (.81)	3.50 (1.73)	3.17 (1.90)	2.96 (1.64)	2.30 (1.72)
Low OTR	.82 (1.17)	.64 (1.03)	1.58 (1.88)	1.17 (1.64)	1.22 (1.59)	.91 (1.38)

number of words retained at 1 and 3 weeks during each condition for the two participant groups and the participants as a whole. These data were used to calculate the effect sizes.

For all participants, the magnitude of effect when comparing the means for the two retention measures was small to medium for each condition with effect size coefficients of $d = .34$, $d = .45$, and $d = .21$ for the high, medium, and low OTR conditions, respectively. Thus, the number of words forgotten across time was relatively small for each of the three conditions. For the average IQ group, effect sizes for word remembering across all conditions were small with coefficients of $d = .18$, $d = .18$, and $d = .23$ for the high, medium, and low OTR conditions, respectively. Thus, although the pairwise comparisons for the average IQ group identified statistically significant mean differences between the 1- and 3-week probes, a relatively small magnitude of forgetting across time during each of the three conditions was demonstrated. This was true even for the high OTR condition where significant forgetting was detected using a paired-samples *t* test. However, for the low IQ group, effect size coefficients of $d = .62$, $d = .80$ were obtained for the high and medium OTR conditions, respectively, indicating a medium to

high magnitude of forgetting across time during these conditions. An effect size coefficient of d = .16 was obtained during the low OTR condition. Although this demonstrates a small magnitude of forgetting across time, students in the low IQ group retained an average of only .82 (SD = 1.17) words at 1 week and retained an average of only .64 (SD = 1.03) words at 3 weeks, leaving little room for forgetting across time. Table 8 displays the effect size coefficients for each group and the participants as a whole.

In reference to the research question, students demonstrated the least forgetting across time when comparing 1- and 3-week retention measures during the low OTR condition. However, conclusions regarding remembering across time during the low OTR condition should be considered in conjunction with the findings of research question number one which found that students retained the fewest number of words during this condition. During the high and medium OTR conditions, significant forgetting did occur. However, the rate of forgetting was very similar, suggesting no difference in word recall across time during these conditions. Perhaps the most interesting finding of the previous analyses was the size of the differences found in remembering across time for the two participant groups. Although pairwise comparisons for both participant groups demonstrated statistically significant differences between the two retention measures, effect size estimates suggested that this difference was larger for the low IQ group. The participants in the average IQ group demonstrated relatively small differences in word recall across time for all three instructional conditions. Conversely, the participants in the low IQ group demonstrated medium to large differences in word recall across time for the

Table 8

Cohen's d *Effect Size Coefficients – 1 to 3 Week Remembering*

	Low IQ	Average IQ	All Participants
High OTR	$d = .62$	$d = .18$	$d = .34$
Medium OTR	$d = .80$	$d = .18$	$d = .45$
Low OTR	$d = .16$	$d = .23$	$d = .21$

high and medium OTR conditions, suggesting that the forgetting magnitude for these students was larger than for the average IQ group.

Research Question Four

Which instructional condition is associated with the least forgetting across time per minute of instruction when comparing 1-week retention measures to 3-week retention measures? It was hypothesized that the medium OTR condition would be associated with the least forgetting across time per minute of instruction when comparing 1- and 3-week measures.

A within-subjects repeated measures ANOVA was performed using the number of words initially learned, retained at 1 week, and retained at 3 weeks per minute of instruction. This analysis indicated a significant main effect for time $F(1.25, 26.31) = 155.3$, $p < .001$, $\eta_p^2 = .88$. Pairwise comparisons were evaluated and revealed significant differences in the number of words initially learned, retained at 1 week, and retained at 3 weeks for each comparison. Specifically, students initially learned more words per minute of instruction than they retained per minute of instruction at 1 and 3 weeks (LSD, $p < .001$ and LSD, $p < .001$, respectively). These findings were expected since efficiency,

when calculated based on initial learning, demonstrates considerably higher values than when it is calculated using retention data (Burns & Sterling-Turner, 2010). Interestingly though, students retained significantly more words per minute of instruction based on the 1-week retention measures than they did based on the 3-week retention measures (LSD, $p = .003$). Although these findings do not demonstrate during which condition the most significant differences across time occurred, or if there were differences between the two participant groups, it does indicate significant differences in the number of words initially learned, retained at 1 week, and retained at 3 weeks per minute of instruction overall.

A significant two-way interaction between condition and time $F(1.94, 40.65) = 47.06$, $p < .001$, $\eta_p^2 = .69$ was found, suggesting that efficiency at each time measurement is related to the instructional conditions, or the number of OTR found within the conditions. Figure 5 depicts this interaction graphically.

Based on a visual inspection of these data, the mean number of words retained per minute of instruction at 1 and 3 weeks during the high and low OTR conditions was nearly identical. Also, the mean number of words retained per minute of instruction at 1 and 3 weeks during the medium OTR conditions was higher than the high and low OTR conditions. In order to determine if the number of words retained per minute of instruction at 1 and 3 weeks were significantly different, paired-samples *t* tests were conducted for each condition. During the high OTR condition, students retained significantly more words per minute of instruction at 1 week than they did at 3 weeks, $t(22) = 3.52$, $p < .05$. Similar findings were demonstrated during the medium OTR condition, $t(22) = 3.07$, $p < .05$. However, during the low OTR condition, students did

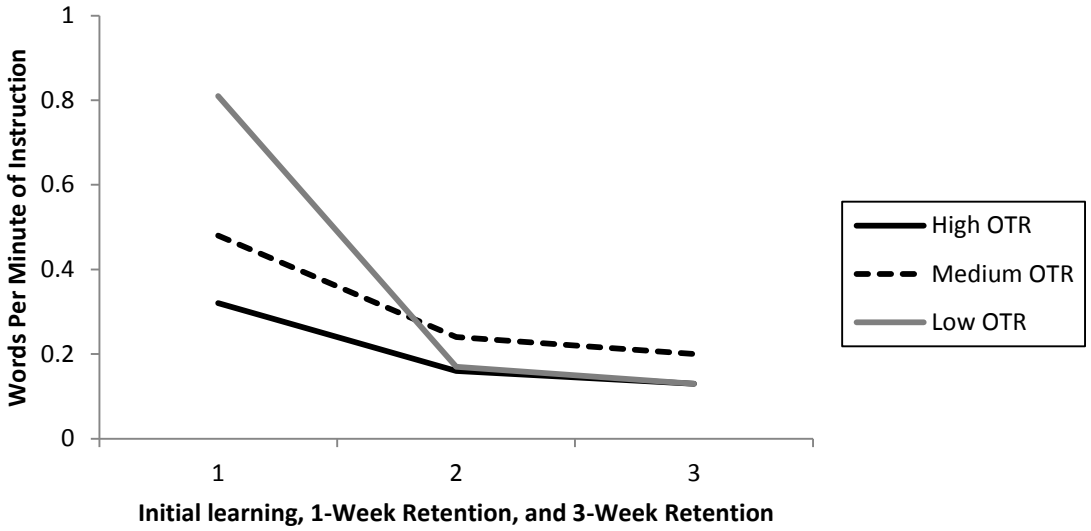

Figure 5. Two-way interaction between condition and time – Efficiency.

not retain a significantly different number of words per minute of instruction when considering the 1- and 3-week probes, t (22) = 1.36, p = .19.

A nonsignificant two-way interaction between group and time F(1.25, 26.31) = .47, p = .54, η_p^2 = .02 was found, suggesting that the words retained at 1 and 3 weeks per minute of instruction for the two participant groups was not statistically different. Therefore, a separate analysis for the two participant groups was not conducted.

In reference to the research question, the least forgetting across time per minute of instruction when comparing 1- and 3-week retention measures occurred during the low OTR condition. However, conclusions regarding these findings should be considered in conjunction with the findings of research question number one which found that students retained the fewest number of words during this condition. Therefore, although students forgot the fewest words per minute of instruction during the low OTR condition, this was likely due to the very low retention rates leaving such a small opportunity to forget words between the 1- and 3-week probes. Also, although students demonstrated significantly

different 1- and 3-week retention per minute of instruction during the high and medium OTR conditions, these findings were not significantly different across the two participant groups.

Research Question Five

Do participants with average IQ scores retain significantly more words (based on 3-week retention measures) than participants with below average IQ scores, as measured at each instructional condition? It was hypothesized that the participants in the average IQ group would retain significantly more words than the participants in the low IQ group during the low and medium OTR condition. However, it was hypothesized that there would not be a significant difference in 3-week retention for the two groups during the high OTR condition.

A within-subjects repeated measures ANOVA was performed using the number of words initially learned, retained at 1 week, and retained at 3 weeks. A significant between-groups effect $F(1, 21) = 8.51$, $p = .008$, $\eta_p^2 = .29$ was found indicating a significant difference in word recall for the two participant groups. However, a nonsignificant interaction between condition and group $F(2, 42) = 1.02$, $p = .37$, $\eta_p^2 = .05$ was found suggesting that the difference between the two groups was not dependent on the conditions. Thus, a similar difference in word recall was found across all conditions. This means that although the low IQ group recalled significantly fewer words than the average IQ group, the difference in word recall across the two groups remained relatively consistent during each condition. This was especially true for the high and medium OTR conditions, which were previously found to be equally effective. These findings are

depicted in Figure 6. A nonsignificant three-way interaction between condition, time, and group was also found $F(4, 84) = 1.25$, $p = .297$, $\eta_p^2 = .06$.

Independent-samples *t* tests were computed to evaluate the difference between the two groups at each condition, during the 1- and 3-week retention measures. During the high, medium, and low OTR conditions, respectively, significant differences in 1-week retention were not found between the groups: $t(21) = 1.79$, $p = .09$; $t(21) = 1.74$, $p = .10$; and $t(21) = 1.16$, $p = .26$. However, during the high and medium OTR conditions, respectively, significant differences in 3-week retention were found between the groups: $t(21) = 3.06$, $p < .05$, and $t(15.14) = 3.01$, $p < .05$. During the low OTR condition, significant differences in 3-week retention were not, however, found between the groups $t(21) = .92$, $p = .37$. These findings further support the findings of research question number three which demonstrated significant forgetting between the 1- and 3-week probes for the low IQ group.

In order to evaluate the magnitude of difference between the two participant groups, Cohen's *d* (Cohen, 1988) was calculated for each condition. Table 9 displays the mean number of words retained for the two participant groups. These data were used to calculate the effect sizes.

For the 1-week retention measures, medium sized differences between the two groups were found with effect size coefficients of $d = .75$ and $d = .73$ reported for the high and medium OTR conditions, respectively. For the 3-week retention measures, much larger differences between the two groups were found with effect size coefficients of $d = 1.29$ and $d = 1.24$ reported for the high and medium OTR conditions, respectively. Medium to small effect size coefficients were found at 1 week ($d = .49$) and 3 weeks ($d =$

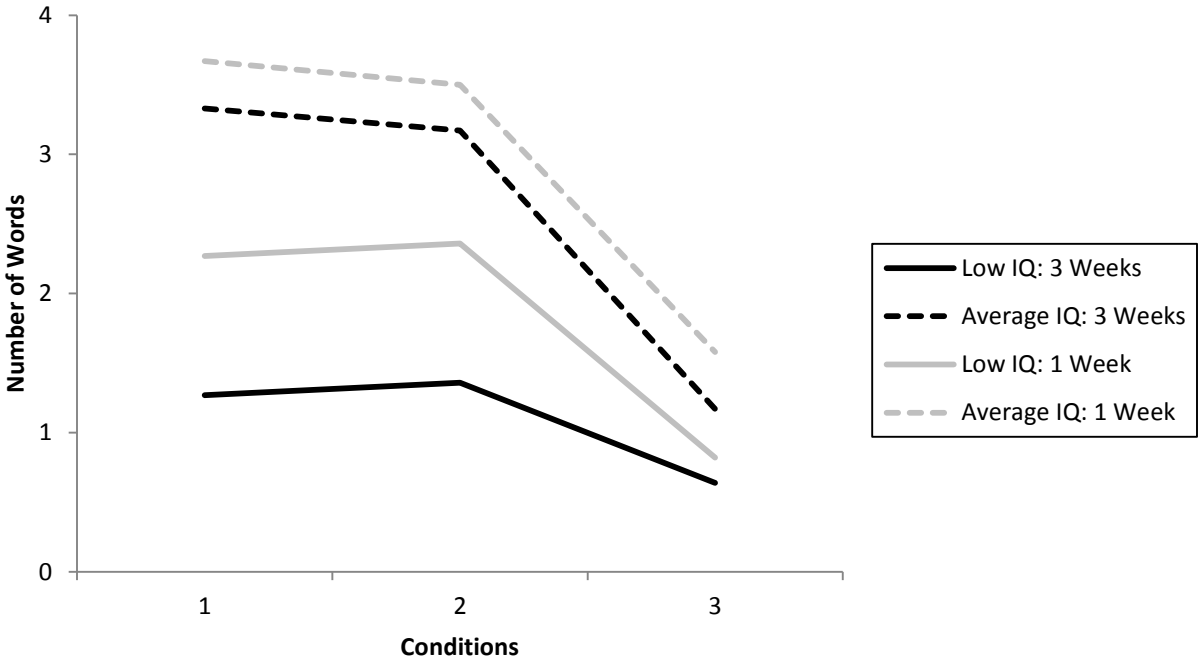

Figure 6. Number of words retained at 1 and 3 weeks for the average and low IQ groups.

Table 9

Number of Words Retained at 1 and 3 Weeks

	Low IQ Group		Average IQ Group	
	1 Week	3 Weeks	1 Week	3 Weeks
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
High OTR	2.27 (1.95)	1.27 (1.19)	3.67 (1.78)	3.33 (1.92)
Medium OTR	2.36 (1.36)	1.36 (.81)	3.50 (1.73)	3.17 (1.90)
Low OTR	.82 (1.17)	.64 (1.03)	1.58 (1.88)	1.17 (1.64)

.39) during the low OTR condition, suggesting smaller differences between the two groups during this condition. Table 10 displays the effect size coefficients for 1- and 3-week retention during each condition.

In reference to the research question, the average IQ group retained significantly more words than the low IQ group during the high and medium OTR conditions when considering the 3-week retention measures. This was in contrast to the initial hypothesis which stated that the average number of words retained by the two participant groups would be the same during the high OTR condition. The average number of words retained at 3 weeks during the low OTR condition was not significantly different for the two groups. Interestingly though, the average number of words retained by the two groups during the 1-week retention measures was not significantly different during any of the conditions. Therefore, contrasting results were found depending on which measure of recall was evaluated.

Research Question Six

Do participants with average IQ scores retain significantly more words per minute of instruction (based on 3-week retention measures) than participants with below average IQ scores, as measured at each instructional condition? It was hypothesized that the average IQ group would retain significantly more words per minute of instruction than the average IQ group across all instructional conditions.

A within-subjects repeated measures ANOVA was performed using the number of words initially learned, retained at 1 week, and retained at 3 weeks per minute of instruction. A significant between-groups effect $F(1, 21) = 28.55$, $p < .001$, $\eta_p^2 = .58$ was found indicating a significant difference in words recalled per minute of instruction for

Table 10

Cohen's d *Effect Sizes – Difference Between Low IQ Group and Average IQ Group*

	High OTR	Medium OTR	Low OTR
1-Week Retention	$d = .75$	$d = .73$	$d = .49$
3-Week Retention	$d = 1.29$	$d = 1.24$	$d = .39$

the two participant groups. However, a nonsignificant interaction between condition and group $F(1.56, 32,85) = .03$, $p = .38$, $\eta_p^2 = .04$ was found suggesting that the difference between the two groups was not dependent on the conditions. Thus, a similar difference in word recall per minute of instruction was found across all conditions. This means that although the low IQ group recalled significantly fewer words per minute of instruction than the average IQ group, the difference across the two groups remained relatively consistent during each condition. These findings are depicted in Figure 7. A nonsignificant three-way interaction between condition, time, and group was also found $F(4, 84) = 2.37$, $p = .108$, $\eta_p^2 = .10$.

Independent-samples *t* tests were computed to evaluate the difference between the two groups at each condition, based on the 1- and 3-week retention measures. During the high and medium OTR conditions, respectively, significant differences in the number of words retained at one week per minute of instruction were found between the two participant groups: $t(21) = 3.279$, $p < .05$; $t(16.62) = 3.18$, $p < .05$. However significant differences were not found during the low OTR condition, $t(14.19) = 1.94$, $p = .07$. During the high and medium OTR conditions, respectively, significant differences in the number of words retained at 3 weeks per minute of instruction were found between the

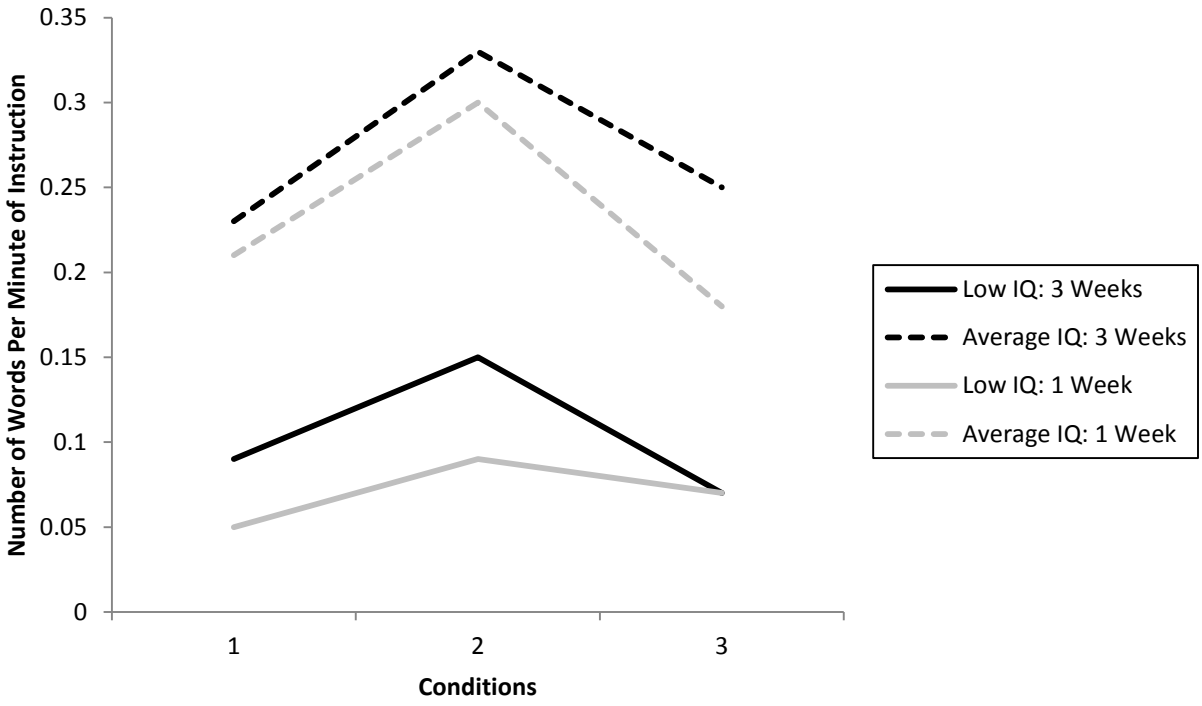

Figure 7. Number of words retained per minute of instruction at 1 and 3 weeks for the average and low IQ groups.

two participant groups: t (13.96) = 4.09, $p < .05$; t (13.28) = 3.74, $p < .05$. However significant differences were not found during the low OTR condition, t (21) = 1.36, p = .20.

In order to evaluate the magnitude of difference between the two participant groups, Cohen's d (Cohen, 1988) was calculated at each condition. Table 11 displays the mean number of words retained for the two participant groups. These data were used to calculate the effect sizes.

For the 1-week retention measures, large sized differences between the two groups were found with effect size coefficients of d = 1.38, d = 1.31, and d = .80 reported for the high, medium and low OTR conditions, respectively. For the 3-week retention

Table 11

Number of Words Retained at 1 and 3 Weeks Per Minute of Instruction

	Low IQ Group		Average IQ Group	
	1 Week	3 Weeks	1 Week	3 Weeks
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
High OTR	.09 (.08)	.05 (.04)	.23 (.11)	.21 (.12)
Medium OTR	.15 (.09)	.09 (.06)	.33 (.17)	.30 (.19)
Low OTR	.07 (.11)	.07 (.10)	.25 (.30)	.18 (.26)

measures, large differences between the two groups were again found with effect size coefficients of $d = 1.68$ and $d = 1.53$ reported for the high and medium OTR conditions, respectively. A medium effect size coefficients was found for 1-week retention during the low OTR condition, $d = .58$. Table 12 displays the effect size coefficients for retention during each condition.

In reference to the research question, the average IQ group retained significantly more words per minute of instruction than the low IQ group during the high and medium OTR conditions. However, the difference between the two groups was not significant during the low OTR condition. This was in contrast to the initial hypothesis which stated that the average number of words retained per minute of instruction by the two participant groups would be the same during all three conditions. Based on Cohen's *d* effect sizes, the differences between the two groups were larger during the high and medium OTR conditions than they were during the low OTR condition.

Table 12

Cohen's d *Effect Sizes – Words Per Minute of Instruction Difference Between Low IQ Group and Average IQ Group*

	High OTR	Medium OTR	Low OTR
1-week Retention	$d = 1.38$	$d = .1.31$	$d = .80$
3-week Retention	$d = 1.68$	$d = 1.53$	$d = .58$

Research Question Seven

Do participants with average IQ scores recall significantly more words than participants with below average IQ scores across the three measurement times (initial learning, 1-week retention measures, and 3-week retention measures)? Participants in each group initially learned six words. It was hypothesized that the participants in the average IQ group would retain significantly more words than the participants in the low IQ group (collapsed across condition) 1 and 3 weeks after instruction.

A repeated measures ANOVA demonstrated a significant between-groups effect $F(1, 21) = 8.51$, $p = .008$, $\eta_p^2 = .29$ suggesting that the high IQ group recalled significantly more words than the low IQ group (collapsed across condition and time). A significant two-way interaction between group and time $F(1.38, 29.04) = 7.29$, $p = .006$, $\eta_p^2 = .26$ was also found, suggesting that the rate of remembering across time for the two participant groups was not the same. All participants initially learned six words during each condition. Averaged across conditions, the average IQ group retained 2.92 (SD = 1.99) words at 1 week and retained 2.56 words (SD = 2.03) at 3 weeks. Averaged across conditions, the low IQ group retained 1.82 (SD = 1.65) words at 1 week and retained 1.09

(*SD* = 1.04) words at 3 weeks. Based on 1-week retention measures, an independent-samples *t* test found that the average IQ group retained significantly more words at 3 weeks (collapsed across condition) than the low IQ group, $t(21) = 2.22$, $p < .05$. Likewise, students in the average IQ group retained significantly more words (collapsed across condition) than the students in the low IQ group, $t(15.42) = 3.64$, $p < .05$. In reference to the research question, participants in the average IQ group recalled significantly more words than the participants in the low IQ group. Figure 8 depicts the group by time interaction.

Research Question Eight

W*hich instructional condition is most preferred by students?* It was hypothesized that the participants would prefer the low OTR condition because it took the least amount of time.

At the conclusion of the study, each student participated in a social validity interview with the experimenter. When asked which condition they preferred, 39% (*n* = 9) chose the high OTR condition, 22% (*n* = 5) chose the medium OTR condition, and 39% (*n* = 9) chose the low OTR condition. Therefore, across all participants, students equally preferred the high and low OTR conditions. However, when considering the participant groups separately, treatment preference was somewhat different. For the average IQ group, 50% (*n* = 6) chose the high OTR condition, 17% (*n* = 2) chose the medium OTR condition, and 33% (*n* = 4) chose the low OTR condition. For the low IQ group, 27% (*n* = 3) chose the high OTR condition, 27% (*n* = 3) chose the medium OTR condition, and 46% (*n* = 5) chose the low OTR condition. Condition preference is outlined in Table 13.

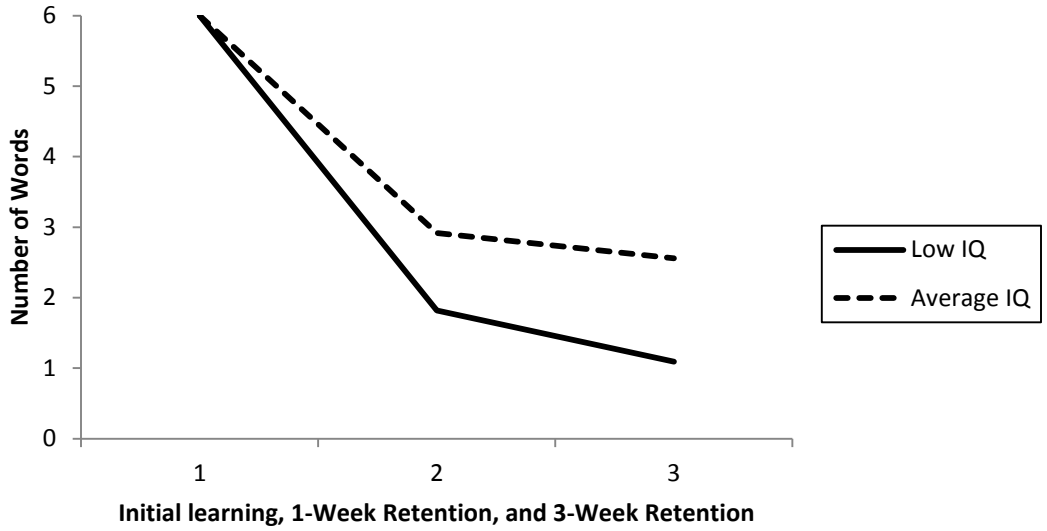

Figure 8. Two-way interaction between group and time – Effectiveness.

Table 13

Condition Preference

	High OTR % (n)	Medium OTR % (n)	Low OTR % (n)
All Participants	39% (9)	22% (5)	39% (9)
Average IQ Group	50% (6)	17% (2)	33% (4)
Low IQ group	27% (3)	27% (3)	46% (5)

Additionally, 83% (*n* = 19) of the overall students felt that the computer program helped them learn new words from the Esperanto language and 87% (*n* = 20) indicated that they enjoyed learning the new words from the Esperanto language. For the average IQ group, 75% (*n* = 9) felt that the computer program helped them learn new words from the Esperanto language and 92% (*n* = 11) indicated that they enjoyed learning the new words from the Esperanto language. For the low IQ group, 91% (*n* = 10) felt that the

computer program helped them learn new words from the Esperanto language and 82% (n = 9) indicated that they enjoyed learning the new words from the Esperanto language.

CHAPTER 4

DISCUSSION

This chapter contains four major sections. The first section includes a discussion of the major findings of this study, as related to previous research. Next, implications for future research are discussed. Third, implications for practice are proposed. Finally, limitations of the study are considered.

Major Findings

Four major findings emerged from this study. First, the high and medium OTR conditions were equally effective and yielded significantly higher retention than the low OTR condition. The fact that students retained a similar quantity of words when they experienced 44 or 27 OTR to each word suggests that there was little benefit of using an IR procedure with more than six known words and 27 OTR to each word. Second, students retained significantly more words per minute of instruction during the medium OTR condition than they did during the high OTR condition. Therefore, although the high and medium OTR conditions were equally effective, the medium OTR condition was significantly more time efficient. Third, based on the 1-week retention measures, the number of words retained by the low and average IQ groups during the high and medium OTR conditions was not significantly different. However, based on the 3-week retention measures, students with low intelligence retained significantly fewer words than students

with average intelligence during the high and medium OTR conditions. Therefore, the students with low intelligence demonstrated a larger magnitude of forgetting across time. Fourth, the largest percentage of students with average intelligence preferred the high OTR condition, even though it took the most time. However, students with low intelligence demonstrated variable preference. Each of these major findings is outlined in detail next.

One of the main purposes of this study was to determine the difference in word recall for three variations of IR using three different levels of OTR, while still retaining the traditional sequence and folding-in component of IR. Overall, the high and medium OTR conditions resulted in an equal number of words retained. Thus, responding to a word 27 times resulted in the same retention as responding to a word 44 times. However, when students responded to words only 14 times during the low OTR condition, their retention was significantly lower. These results are in alignment with the findings of Szadokierski and Burns (2008) who demonstrated nonsignificant differences in 1-week retention between words with 18 to 60 OTR. However, as with the current study, Szadokierski and Burns found that when students experienced fewer than 18 OTR, their word retention significantly declined. As described by Ebbinghaus (1885), the learning curve demonstrated in this study was similar to that found by Szadokierski and Burns. The number of words retained reached a plateau at 27 and 44 OTR. Thus, results of this study are consistent with the findings of Szadokierski and Burns who suggested that there is little benefit to providing more than 18 response opportunities.

Although Szadokierski and Burns (2008) effectively demonstrated that OTR was more highly correlated with retention than drill ratio, the current study extends the

findings of Burns (2004b) who found no effectual differences when using eight or nine known words. This study demonstrated that using six known words was just as effective as using eight known words within the IR procedure. Again, as identified by Szadokierski and Burns (2008), OTR is significantly more correlated with retention than drill ratio, so the drill ratios are only important in this case because they identified the number of OTR found within each condition. In this study, six known words equated to 27 OTR to each unknown word and eight known words equated to 44 OTR to each unknown word. Thus, using nine known words as originally outlined by Tucker (1989) may not be necessary.

As suggested by Szadokierski and Burns (2008), the current study also examined the effect of OTR on 1- and 3-week recall. Based on both the 1- and 3-week retention measures, there were no significant differences found between the high and medium OTR conditions. Additionally, word retention was consistent across the high and medium OTR conditions for the two participant groups. Thus, even with students with low intelligence, there were no differences in the number of words retained at 1 and 3 weeks when students experienced 27 or 44 OTR.

The previous findings regarding the effectiveness of the high and medium OTR conditions are directly related to the efficiency of the two conditions. As originally outlined by Tucker (1989), IR intersperses one unknown word among nine known words. Although using nine known words facilitates high OTR, the current study demonstrated that this many OTR may not be necessary and causes IR to be less time efficient. Students learned the same number of words during the high and medium OTR conditions, yet the medium OTR condition took less time and was significantly more time efficient than the high OTR condition.

The efficiency of each instructional condition was evaluated by dividing the number of words initially learned or retained at 1 or 3 weeks by the duration of the instructional session in minutes. On average, the duration of each instructional session was about six minutes different from each other with the high, medium, and low OTR conditions lasting 19.98 minutes (*SD* = 5.96), 13.86 minutes (*SD* = 4.73), and 8.13 minutes (*SD* = 2.82), respectively. Thus, the medium OTR took about six fewer minutes than the high OTR condition, but the same number of words were recalled. It is important to note that these results were found for both the low and average IQ groups.

Despite the preponderance of literature supporting the effectiveness of IR, it has often been criticized because it takes more instructional time than other drill and practice procedures (Nist & Joseph, 2008; Skinner, 2008). However, these criticisms are based on efficiency calculated using the number of words initially learned or learned per minute of instruction rather than the number of words retained per minute of instruction (Burns & Sterling-Turner, 2010). Recent studies have found that when based on 1- or 2-week retention measures, the efficiency of IR is equivalent to, or better than other drill methods (Burns & Boice, 2009; Burns & Sterling-Turner, 2010). This study supports the findings of Burns and Boice (2009) and Burns and Sterling-Turner (2010) regarding the metric used to calculate efficiency. For example, students in the current study learned .81 words per minute of instruction during the low OTR condition and .32 words per minute of instruction during the high OTR condition. Therefore, when calculated based on the number of words initially learned, the low OTR condition was clearly more efficient. However, when calculated using retention measures, students retained the same number of words per minute of instruction during the high and low OTR conditions. Based on 1-

week retention measures, students retained .17 words per minute of instruction during the low OTR condition and .16 words per minute of instruction during the high OTR condition. When calculated using the 3-week retention measures, students retained .13 words per minute of instruction during both the high and low OTR conditions. To put this into perspective, students retained an average of 2.35 (*SD* =1.90) words during the high OTR condition and only .91 (*SD* = 1.38) words during the low OTR condition. Thus, although the low and high OTR conditions appeared to be equally efficient based on the 3-week retention measures, significantly more words were retained during the high OTR condition than during the low OTR condition. Therefore, it is important to consider effectiveness and efficiency when evaluating the overall value of an instructional intervention. Although the current study did not compare IR to any other drill and practice procedures, it did demonstrate that using only six known words within the IR procedure was equally effective but more efficient than using eight known words. In this study, the efficiency of IR was improved, without jeopardizing its effectiveness.

Students in the low IQ group retained significantly fewer words per minute of instruction than the students in the average IQ group during both the high and medium OTR conditions. However, the number of words initially learned, retained at 1 week, and retained at 3 weeks per minute of instruction followed a similar trend for the two participant groups (see Figure 7), with the medium OTR condition being the most time efficient for both groups. Since students demonstrated the same recall during the high and medium OTR conditions, but the medium OTR condition took less time because fewer known words were rehearsed, these results were expected.

Although the mean number of words retained at 1 week by students in the low IQ group was lower than the mean number of words retained at 1 week by students in the average IQ group, the difference between the two groups was not significantly different during the high or medium OTR conditions. This is an important finding and supports the assertion of previous researchers who suggested that IR could negate the effect of individual differences such as intelligence or verbal ability (Burns & Boice, 2009; MacQuarrie, 2002). However, based on 3-week retention measures, the number of words retained during the high and medium OTR conditions was significantly different for the two groups. The participants in the low IQ group forgot a large number of words between the 1- and 3-week retention measures during the high (d = .62) and low (d = .80) OTR conditions, whereas the average IQ group did not (d = .18; d = .18).

These findings are dissimilar to the findings of Burns and Boice (2009) who also taught Esperanto words using IR to seventh and eighth graders with low intelligence. These students had an average IQ of 74.40 (SD = 9.79). They were taught nine words during the IR condition and remembered an average of 5.30 (SD = 2.30) words 1 week later and an average of 4.95 (SD = 2.21) words 2 weeks later. Therefore, at 1 and 2 weeks, these students recalled an average of 59% and 55% of the words initially taught, respectively, with no indication of significant forgetting over time. In the current study, students were taught six Esperanto words during both the high and medium OTR conditions. During the high OTR condition, students recalled an average of 2.27 (SD = 1.95) at 1 week and an average of 1.27 (SD = 1.19) at 3 weeks. During the medium OTR condition, students recalled an average of 2.36 (SD = 1.36) at 1 week and an average of 1.36 (SD = .81) at 3 weeks. Therefore, at 1 and 3 weeks, students retained 38% and 21%

of the words during the high OTR condition and 39% and 23% of the words during the medium OTR condition.

The current study demonstrated two major findings that were different from the findings of Burns and Boice (2009). First, students in the current study recalled a smaller percentage of words 1 week after instruction than students did in the study by Burns and Boice. Second, the difference between the two measures of recall was much larger for students in the current study (High OTR, d = .62; Medium OTR, d = .80) than they were for students in the study by Burns and Boice (d = .16). Although the magnitude of this difference, as demonstrated by Cohen's *d,* was not reported in the study by Burns and Boice, it is reported here and was calculated based on the means and standard deviations provided by the authors. These findings demonstrate that forgetting across time was much higher for the students with low IQ in the current study. Although the mean IQ score for students in the study by Burns and Boice was only 6 points higher than the mean IQ score for the students in the current study, this could be a possible explanation for the difference in forgetting across time found in the two studies. Also, Burns and Boice measured retention at 1 and 2 weeks, whereas the current study measured retention at 1 and 3 weeks. Therefore, the additional time delay between recall measures in the current study may account for some of these differences.

The study by Burns and Boice (2009) was a replication of a study by MacQuarrie et al. (2002). However, MacQuarrie et al. used participants with a mean standard score of 101.14 (11.69) on the Peabody Picture Vocabulary Test-Revised (PPVT-R). Seventh-graders were taught nine Esperanto words during the IR condition and remembered an average of 6.16 (SD = 1.60) words 1 week later and an average of 5.17 (SD = 1.85) words

30 days later. Therefore, at 1 week and 30 days, these students recalled an average of 68% and 57% of the words initially taught, respectively. In the current study, students were taught six Esperanto words during both the high and medium OTR conditions. During the high OTR condition, students in the average IQ group retained an average of 3.67 (*SD* = 1.78) at 1 week and an average of 3.33 (*SD* = 1.92) at 3 weeks. During the medium OTR condition, students retained an average of 3.50 (*SD* = 1.73) at 1 week and an average of 3.17 (*SD* = 1.90) at 3 weeks. Therefore, at 1 and 3 weeks, respectively, students retained 61% and 56% of the words during the high OTR condition and 58% and 53% of the words during the medium OTR condition. The findings of MacQuarrie et al. are in alignment with the findings of the current study for students with average IQ.

By using intelligence as a between-groups factor, this study demonstrated that the number of words retained 1 week after instruction by students with average IQ and low IQ was not significantly different. However, the number of words retained by the two groups 3 weeks after instruction was significantly different. Thus, the current study supports the assertion of previous researchers regarding the possibility of IR negating the effect of individual differences such as IQ (Burns & Boice, 2009; MacQuarrie et al., 2002). However, 3-week retention for the low IQ group was lower than for students with average intelligence. Therefore, students with low IQ may need additional and intermittent instruction or practice opportunities to avoid forgetting the information they have learned using IR. These findings have implications for future research and practice.

Although the low IQ group forgot a large number of words between the two retention measures, the rate of forgetting was consistent across the high and medium OTR conditions. Therefore, as discussed previously, the consistency of forgetting

between the high and medium OTR conditions further demonstrates the similarity of these two conditions.

Results regarding treatment preference were mixed when examining the treatment acceptability forms for the participants as a whole. However, when examining the results for each group individually, there were some slight differences. Exactly half of the students in the average IQ group indicated that they preferred the high OTR condition. When asked why, five of the six students who preferred this condition indicated that practicing words more times during this condition helped them to have better recall. Interestingly though, only 17% of the students in the average IQ group preferred the medium OTR condition, even though recall during these two conditions was equivalent. Thirty-three percent of the students in the average IQ group preferred the low OTR condition. Three of the four students who preferred this condition liked that it took less time and felt that the other conditions were too long. This explanation was similar to that of the students in the study by Nist and Joseph (2008) who preferred the traditional drill and practice condition to the IR condition because it took less time. Thus, while some students felt that the time spent learning words during the high OTR condition was beneficial, other students preferred to complete the instructional session more quickly, even though word recall was negatively impacted.

The largest percentage of students in the low IQ group preferred the low OTR condition (46%), with equal amounts of students preferring the high and medium OTR conditions (27%, 27%). When asked why a particular condition was preferred, the responses from the students in the low IQ group were extremely variable and typically were not related to the intervention. For example, 5 of the 11 participants in the low IQ

group preferred a particular condition because their favorite color was displayed as the background color on the PowerPoint presentation. With the exception of 4 participants, none of the reasons students identified were related to the intervention. Two students preferred the high OTR condition because they felt seeing the words more frequently helped them learn the words better, and two participants preferred the low OTR condition because it was the shortest.

When asked which condition they preferred, many of the students in both participant groups either looked confused or did not know the difference between the conditions. In these situations, the experimenter briefly explained to the students that they saw each word 44 times during the condition with the blue background computer screen, and that this condition took the longest amount of time. Students were also told that the condition with the yellow background presented each word 27 times, but this took less time than the blue one, and that the condition with the red screen presented the words 14 times, and this condition took the least amount of time. Therefore, it is possible that prompting from the experimenter influenced some of the student's responses.

Implications for Future Research

The current study compared the effectiveness and efficiency of three variations of IR. While the OTR to each word was different in each of the three conditions of this study, this was the first study to truly hold the number of OTR to each word within the IR procedure constant, while still retaining the folding-in component and providing a high number of response opportunities. This was achieved by re-introducing the known words as unknown words in order to help move the targeted unknown words all the way through the IR procedure. While the practice of re-introducing known words may not be practical

in applied settings because it increases the session length, future researchers may wish to use this methodology in order to objectively compare IR to other drill procedures, while truly holding OTR constant. For example, since the current study demonstrated no difference between the high and medium OTR conditions, future research could compare an IR procedure with six known words, and 27 OTR to each unknown word, to a traditional drill procedure where students also respond to each word 27 times. Burns and Sterling-Turner (2010) found an IR procedure using eight known words to be equally efficient as a traditional drill and practice procedure when calculated using 1-week retention measures. Since the IR procedure using six known words in the current study was equally effective and more efficient than the IR procedure using eight known words, it is possible that IR using only six unknown words would be more effective and time efficient than a traditional drill and practice procedure with the same number of OTR to each word.

The order in which words are presented and folded-in within the IR procedure may be related to the effectiveness of IR. Future researchers could examine the relationship between word presentation order and word retention to determine if words presented near the beginning or end of the procedure are recalled more frequently than others. As traditionally implemented, words introduced towards the end of the IR session are not folded-in as known words because the procedure ends at this point. Thus, it is possible that since the OTR to these words is so low, retention would be negatively impacted. Also, since students demonstrate better long-term retention of words taught using IR than they do for words taught using other drill procedures (Burns & Boice, 2009; Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002), long-term retention

should continue to be evaluated in future research. A large portion of research using IR has used Esperanto words. While this may increase internal validity, using actual words increases external validity. Future research could replicate this study using grade appropriate English sight words and measure generalization.

Words were presented in the current study on a laptop computer using a Microsoft PowerPoint presentation. Not only did this procedure reduce the chance for procedural error, experimenters were able to quickly advance to the next word without the burden of handling the flash cards and keeping them in the correct sequence. Although the current study did not directly test this hypothesis, it is possible that using a computer, rather than typical 3x5 flash cards, reduces the duration of instructional sessions because interventionists do not need to arrange or handle flash cards. Future research could evaluate the effectiveness and efficiency of a computer-based IR procedure in comparison to an IR procedure using index cards. Additionally, with the increased interest and use of tablet computers such as the iPad and other touch screen devices in school settings, future researchers could implement IR using these modalities. Very favorable treatment acceptability was demonstrated using an iPod to deliver a time-delay taped-words procedure (e.g., Pummel, 2011), thus research using similar technology to implement IR seems warranted. Many classrooms are now equipped with interactive touch sensitive white boards such as the SMART Board. Using this technology, it is possible that IR could be used as a classroom center, with students acting as peer tutors, allowing time for the teacher to provide direct instruction to other students. Or, as an alternative, the classroom teacher could direct a choral responding activity with a small

group or whole class using IR on a SMART Board. Future research evaluating the effectiveness of IR using peer tutoring or choral responding is needed.

The current study found that students with low intelligence forgot significantly more words over time than students with average intelligence. Future research could examine the effect of providing additional response opportunities intermittently between instruction and recall probes for students with low intelligence to see if this improves their long-term retention. Additionally, since all participants, including the students in the average IQ group forgot half or more of the words they initially learned, research examining a distributed practice variation of IR could be conducted. Providing practice opportunities more frequently may lead to better retention. Research could examine the effect of spreading OTR across several days rather than providing a large number of OTR within a single instructional session.

Implications for Practice

Results of the current study demonstrated several practical applications for applied settings. First, an IR procedure using six known words (27 OTR to each unknown word) was equally effective as an IR procedure using eight known words (44 OTR to each unknown word). Although using fewer known words resulted in fewer response opportunities to the unknown words, this study demonstrated no difference in recall when students were given 27 OTR compared to 44 OTR. Also, reducing the number of known words and OTR to each word caused the IR procedure to be more time efficient. Practitioners could consider using this variation of IR in practice since it was equally effective as a traditional IR procedure using more known words, but took less instructional time.

As recommended by Burns and Sterling-Turner (2010), this study also found value in using retention data when determining the efficiency of an intervention. When based on initial learning, the low OTR condition was the most efficient intervention. However, when based on retention measures, the medium OTR condition was the most time efficient and the high and low OTR conditions were equally efficient. Therefore, when selecting interventions, teachers should consider instructional efficiency based on retention, in addition to initial learning. An instructional procedure that yields efficient learning, but only moderate retention may not be as desirable as a procedure that requires more instructional time but facilitates high levels of retention.

The findings of this study for students with low intelligence also have significant implications for practice. The high and medium OTR conditions were equally effective for this group of students as well, with the medium OTR condition being the most time efficient. Additionally, based on 1-week retention measures, the number of words retained by students with low intelligence was not significantly different than the number of words retained by students with average intelligence. As suggested by previous researchers (Burns & Boice, 2009; MacQuarrie et al., 2002), this study provides support to the claim that IR may negate the effects of individual differences such as intelligence. Thus, teachers can confidently use IR with students with low intelligence. In comparison to students with average intelligence, however, these students may require intermittent additional instruction or practice opportunities after the initial learning due to their weaker long-term retention. As conducted in this study, IR was used as a mass practice procedure. However, teachers may want to use a distributed practice model by teaching only a few words at a time during shorter sessions. Then, the IR session can begin where

it was left off, with the previously taught words still being practiced as folded-in known words. This may be beneficial for students with low and average intelligence since both groups of students in the current study forgot a significant number of words 1 week after initially learning them.

This study presented words using a Microsoft PowerPoint presentation on a laptop computer rather than using index cards. Overall, students found the intervention to be favorable. Previous research has also found high treatment acceptability ratings for instructional practices using technology (Pummel, 2011). Although not empirically evaluated, IR may have practical applications in classrooms using interactive touch screen white boards such as the SMART board or tablet computers such as the iPad.

Limitations

Several limitations regarding the external and internal validity of this study should be pointed out. First, unknown words were selected from the Esperanto international language. Although this increased the internal validity of the findings, it is unknown if similar results would be found had English words been used. Second, this study utilized the previously untested procedure of folding-in known words until all target unknown words had completely cycled through the IR sequence. Although this procedure took additional instruction time, it allowed each word to be practiced the same number of times within a given condition. The additional time spent rehearsing known words may make it difficult to compare the efficiency results of this study to other studies that did not use this procedure. Additionally, this may not a reasonable procedure to use in practice because valuable instructional time is used rehearsing words the student already knows. However, within a research setting, it seems like a reasonable method of holding

the variable of OTR constant in order to compare two interventions. Third, since middle school students were used in this study, it is unknown if the results will generalize to younger students. Fourth, error correction procedures during the instructional sessions were needed less than three times per word for the majority of the students in this study. However, two students in the low IQ group made a significant number of errors on multiple words, demonstrating poor acquisition. As described by Burns (2004a), if students make more than three errors while learning a particular word, they have met their acquisition rate and the procedure should end. Consideration to acquisition rate was not included in this study. Finally, no measure of generalization was used so it is unclear if students would have generalized the words they learned. However, one student incidentally commented that she liked learning the Esperanto words because she had been calling her brother a hundo (dog) and a blato (cockroach) without him knowing what she was saying. Thus, apparently some generalization occurred, at least for this student.

A few limitations related to the methodology and measurement procedures of the study also existed. First, the first 24 students with signed permission and assent forms, and who met the IQ inclusion criteria, were included in the study. Although the final participants consisted of a fairly equal distribution of gender, age, and race, the participant sample was not randomly selected. Second, three lists of six Esperanto words were randomly assigned to each condition, creating nine variations of word sets and conditions. The order in which students experienced the conditions were also randomized and counterbalanced. However, the words within each of the sets were presented in the same order to each student, regardless of the condition they taught in. This was done so that a different PowerPoint presentation was not required for each participant during each

condition. Third, due to the similar nature of the three conditions, some students did not know the difference between them during the treatment acceptability interviews. Some responses may have therefore been skewed by the explanation provided by the experimenter.

APPENDIX A

PARENT CONSENT TO REVIEW SPECIAL EDUCATION RECORDS

Special Education Records Review Permission Form

BACKGROUND

This letter is to inform you of a research study being conducted at your child's school. The study will use a sight word instructional procedure called Incremental Rehearsal to teach words to students with below average intelligence who receive special education services, as well as to students with average intelligence who do not receive special education services. This study is being conducted as part of Kade Johnson's doctoral dissertation in educational psychology at the University of Utah. Although Mr. Johnson is currently a practicing school psychologist in the Florence Unified School District (FUSD), this study is not sponsored by FUSD. However, FUSD and the principal at Anthem K-8 have reviewed and approved this study. We are hopeful that this study will help us learn better ways to teach sight words to children in the future.

You are receiving this letter because your child currently receives special education services and may meet the initial participant selection criteria. However, we need your permission to review their special education records in order to determine if they meet the initial participant selection criteria. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you will allow your child's special education records to be reviewed.

PROCEDURE

The research study involves the participation of students who receive special education services and who have below average intelligence, as demonstrated on standardized measures of intelligence. In order to identify potential participants for this study, Mr. Kade Johnson will review Multidisciplinary Evaluation Reports contained within student's special education records. Standardized intelligence scores found within these reports will be extracted and recorded on a spreadsheet with the students name, birth date, and gender. Mr. Johnson will be the only person to review the special education records. However, members of the study team will have access to the spreadsheet where this data is recorded. Students with IQ standard scores ranging from approximately 55 to 77.5 will be considered as potential participants. If your child meets the initial criteria for participation; you may be contacted again to seek permission for them to participate in the actual study. At that point, the study will be explained to you in detail and you will have an opportunity to ask any questions. If your child does not meet the initial participant selection criteria, you will not be contacted again. By signing this form, you are giving consent for your child's most recent Multidisciplinary Evaluation Report to be reviewed. Signing this form does not mean that your child will participate in the study. It only means that you are giving consent for your child's special education records to be reviewed. All information regarding your child will remain confidential.

RISKS

The risks relative to your child's special education records being reviewed are minimal. While the purpose of the records review is to extract intelligence scores, it is possible that

the researcher will view other portions of the special education records while trying to locate the intelligence scores. However, all information will remain confidential.

BENEFITS
There are no anticipated direct benefits to your child as a result of the records review.

CONFIDENTIALITY
Information about your child will be kept confidential. Information extracted from the special education records will be stored in a locked filing cabinet or on a password protected computer located in the researcher's work space. Only the researcher and members of his study team will have access to this information. All personally identifiable information will be destroyed at the conclusion of the study.

PERSON TO CONTACT
If you have questions, complaints or concerns about this study, you can contact Mr. Kade Johnson at 520-723-6400 who may be reached during regular school hours, 8:00 to 4:30. Mr. Johnson can also be reached via email at kadejohnson@fusdaz.org. You may also contact Dr. Amy Fuller, Florence Unified School District Assistant Superintendent, if you have questions or concerns by calling (520) 866-3512 or emailing afuller@fusdaz.org.

Institutional Review Board: Contact the Institutional Review Board (IRB) if you have questions regarding you or your child's rights as a research participant. Also, contact the IRB if you have questions, complaints or concerns which you do not feel you can discuss with the investigator. The University of Utah IRB may be reached by phone at (801) 581-3655 or by e-mail at irb@hsc.utah.edu.

Research Participant Advocate: You may also contact the Research Participant Advocate (RPA) by phone at (801) 581-3803 or by email at participant.advocate@hsc.utah.edu.

VOLUNTARY PARTICIPATION
It is up to you to decide whether to allow your child's special education records to be reviewed. Refusal will involve no penalty or loss of benefits to which your child is otherwise entitled. This will not affect your or your child's relationship with the investigator.

COSTS AND COMPENSATION TO PARTICIPANTS
There is no cost or compensation associated with your child's records being reviewed.

CONSENT
By signing this consent form, I confirm I have read the information in this parental permission form and have had the opportunity to ask questions. I will be given a signed copy of this parental permission form. I voluntarily agree to allow my child's special education records to be reviewed. I understand that I may or may not be contacted in the future to seek permission for my child to participate in the actual study.

_____
Child's Name

_____
Parent/Guardian's Name

_____ _____
Parent/Guardian's Signature Date

_____
Relationship to Child

_____
Name of Person Obtaining Consent

_____ _____
Signature of Person Obtaining Consent Date

APPENDIX B

TEACHER LETTER

December 15, 2011
RE: Kade Johnson's doctoral dissertation research study

Dear Teacher,

As part of my doctoral dissertation in educational psychology at the University of Utah, I will be conducting a research study using students at Anthem K-8 as participants. Although I am currently a practicing school psychologist in the Florence Unified School District (FUSD), this study is not sponsored by FUSD. However, FUSD and the principal at Anthem K-8 have reviewed and approved this study.

Students will be taught the pronunciation and English translation of words from the Esperanto international language using a flashcard technique called incremental rehearsal. Esperanto words, rather than English words, will be taught since the likelihood of previous exposure is very low, and in order to control for possible learning opportunities between the instructional sessions and maintenance measures. Incremental rehearsal is a systematic procedure in which unknown words are interspersed and taught among words the student already knows. Although research has shown incremental rehearsal to be an effective intervention, it takes much longer than traditional flashcard techniques that rehearse only unknown words. Since instructional time is such a valuable resource in schools, using interventions that are both effective and time efficient is important. This research study will compare three variations of incremental rehearsal to determine which one is the most effective and time efficient. Some research suggests that incremental rehearsal can be as effective for students with below average intelligence as it is for students with average intelligence. Thus, the effectiveness and time efficiency of each of the three incremental rehearsal instructional conditions will be compared across students with varying levels of intelligence. A potential pool of participants with below average intelligence will be selected based on a review of special education records. A potential pool of participants with average intelligence will be selected based on teacher nomination. This is where we need your help.

Where indicated on the reverse side of this letter, please recommend three potential participants who display average academic skills with no suspected learning disability or cognitive delay. I will contact parents of these students to explain the purpose of this study, inclusion criteria, study procedures, and potential risks and benefits. Parents and students will be asked to sign consent and assent forms if they choose to participate in the study. As a teacher, you are being asked to recommend potential participants who may or may not be selected for participation in the study. Your recommendation does not imply or require that these students participate.

If selected for participation, this study will require approximately two total hours of time for each participant across a span of six weeks. The first three sessions will take 10-25 minutes, and the remaining sessions will take only a few minutes. On the reverse side of this letter, please indicate several minimally intrusive times, such as silent reading time, recess, home room, etc., for the student to participate in the study sessions. If your student is selected for participation, I will regularly communicate with you to ensure that critical instructional time is not missed.

If you have any questions or concerns, please let me know by calling ext 3016, or emailing kadejohnson@fusdaz.org. You may also contact Dr. Amy Fuller, FUSD

Assistant Superintendent, if you have questions or concerns by calling (520) 866-3512 or emailing afuller@fusdaz.org.

Thank you very much for your help with this process.

Thank you,

Kade Johnson

Teacher Name: _____

Please provide the name and grade of three potential participants below:

Name: _____ Grade: _____

Name: _____ Grade: _____

Name: _____ Grade: _____

Please indicate preferred times for your students to participate in the study sessions, if they are selected.

_____
_____
_____
_____
_____
_____

Please return this form to me within two days of receiving it. Thanks for your help.

APPENDIX C

PARENT PERMISSION

Parental Permission Document

BACKGROUND

Your child is being asked to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether you will allow your child to take part in this study. If you decide to allow your child to participate in this study, they will also have the opportunity to give their assent to participate by signing a child assent form. The primary researcher will read the child assent form to your child and allow them to ask questions. If your child decides they do not want to participate, even if you give your permission, they will not be required to participate. A blank copy of the child assent form is attached to this notice for your information. You will be given a copy of the signed assent form if your child decides to participate.

This study is being conducted as part of Kade Johnson's doctoral dissertation in educational psychology at the University of Utah. Although Mr. Johnson is currently a practicing school psychologist in the Florence Unified School District (FUSD), this study is not sponsored by FUSD. However, FUSD and the principal at Anthem K-8 have reviewed and approved this study.

This study will teach unknown words to students using a flashcard technique called incremental rehearsal. Incremental rehearsal is a systematic procedure in which unknown words are interspersed and taught among words the student already knows. Although incremental rehearsal has been shown to be effective in previous research, it takes much longer than traditional flashcard techniques that rehearse only unknown words. Since instructional time is such a valuable resource in schools, using interventions that are both effective and time efficient is important. This research study will compare three variations of incremental rehearsal to determine which one is the most effective and time efficient. Some research suggests that incremental rehearsal can be equally as effective for students with below average intelligence as it is for students with average intelligence. Thus, the effectiveness and time efficiency of each of the three incremental rehearsal instructional conditions will be compared across students with varying levels of intelligence.

STUDY PROCEDURE

Before final selection for participation in the study, students will be administered the Kaufman Brief Intelligence Test, Second Edition (KBIT-2) to determine if he or she meets the intelligence score inclusion criteria. By signing this form and giving permission for your child to participate in this study, you are also giving permission for the KBIT-2 to be administered to your child. Participants will be selected and included in one of two participant groups: (1) students who are currently receiving special education services and have below average intelligence standard scores ranging from 55 to 77.5, and (2) students who are not currently receiving special education services and have average intelligence standard scores ranging from 92.5 to 115. Each of the potential participants from the below average intelligence group were initially selected based on their

participation in special education programming and a review special education files. The remaining potential participants were selected based on teacher recommendation of students with average academic skills and no suspected learning or cognitive delays. Thus, if you decide to give permission for your child to participate in this study, the KBIT-2 will be administered to them first. If your child meets the intelligence score inclusion criteria, he or she will be selected for participation in the study. You will be notified if your child has, or has not, been selected as a final participant in this study.

If selected for participation, your child will be taught the pronunciation and English translation of words from the Esperanto international language. Esperanto words, rather than English words, will be taught since the likelihood of previous exposure is very low, and in order to control for possible learning opportunities between the instructional sessions and maintenance measures. The use of Esperanto words has been commonly used in previous published studies using incremental rehearsal. A few examples of Esperanto words are stelo (star) and pilko (ball).

Participants will meet one-on-one with an experimenter during six sessions, with each session spaced exactly one week apart. The experimenters will be one of four school psychology graduate students who have been trained to implement the study procedures. During the first three sessions, students will be taught words using three different instructional conditions. For each of these sessions, the words will be presented to students on a computer screen rather than using traditional paper flash cards. Measures of word maintenance will occur one and three weeks after each instructional condition. The instructional conditions will last approximately 10-25 minutes, and maintenance measures will take approximately one minute. Additionally, administration of the KBIT-2 will take approximately 15-30 minutes. Thus, this study will require about two hours or less for each participant across a span of six weeks. Every effort will be made to only pull students from class during non-instructional times such as silent reading time, recess, home room, before and after school, or other minimally intrusive times. The experimenter will communicate with the classroom teacher and student to ensure critical instructional time is not missed. The classroom teacher and student have the right to delay participation until a more convenient time. Once the study is completed, the researcher will evaluate the data to determine which instructional condition was the most effective and time efficient for the two groups of participants.

RISKS

The risks of this study are minimal. Flash card instruction, as well as computer-based instructional procedures, are commonly used in schools and are not considered experimental or risky procedures. Your child may or may not find this experience enjoyable. While students will not be removed from their class during critical instructional times, they will miss a total of approximately two hours of class across a span of six weeks.

BENEFITS

There are no anticipated direct benefits to your child for taking part in this study. However, your child may enjoy the novelty of learning words from the Esperanto

international language. Additionally, we hope the information we get from this study will help develop a greater understanding of sight word instructional procedures in the future.

CONFIDENTIALITY
Information about your child will be kept confidential. Data and records will be stored in a locked filing cabinet or on a password protected computer located in the researcher's work space. Only the researcher and members of his study team will have access to this information.

However, if your child discloses actual or suspected abuse, neglect, or exploitation of a child, or disabled or elderly adult, the researcher or any member of the study staff must, and will, report this to Child Protective Services (CPS), Adult Protective Services (APS) or the nearest law enforcement agency.

PERSON TO CONTACT
If you have questions, complaints or concerns about this study, you can contact Mr. Kade Johnson at 520-723-6400. If you feel your child has been harmed as a result of participation, please contact Mr. Johnson who may be reached during regular school hours, 8:00 to 4:30. Mr. Johnson can also be reached via email at kadejohnson@fusdaz.org. You may also contact Dr. Amy Fuller, Florence Unified School District Assistant Superintendent, if you have questions or concerns by calling (520) 866-3512 or emailing afuller@fusdaz.org.

Institutional Review Board: Contact the Institutional Review Board (IRB) if you have questions regarding your rights as a research participant. Also, contact the IRB if you have questions, complaints or concerns which you do not feel you can discuss with the investigator. The University of Utah IRB may be reached by phone at (801) 581-3655 or by e-mail at irb@hsc.utah.edu.

Research Participant Advocate: You may also contact the Research Participant Advocate (RPA) by phone at (801) 581-3803 or by email at participant.advocate@hsc.utah.edu.

VOLUNTARY PARTICIPATION
It is up to you to decide whether to allow your child to take part in this study. Refusal to allow your child to participate or the decision to withdraw your child from this research will involve no penalty or loss of benefits to which your child is otherwise entitled. This will not affect your or your child's relationship with the investigator.

COSTS AND COMPENSATION TO PARTICIPANTS
There is no cost associated with participation in this study. As compensation, participants will be offered a small piece of candy, small toy, or sticker from a prize box after each session.

CONSENT

By signing this consent form, I confirm I have read the information in this parental permission form and have had the opportunity to ask questions. I will be given a signed copy of this parental permission form. I voluntarily agree to allow my child to take part in this study. I also give permission for the KBIT-2 to be administered to my child.

_____
Child's Name

_____
Parent/Guardian's Name

_____ _____
Parent/Guardian's Signature Date

_____
Relationship to Child

_____
Name of Person Obtaining Consent

_____ _____
Signature of Person Obtaining Consent Date

APPENDIX D

CHILD ASSENT

Assent to Participate in a Research Study

Who are we and what are we doing?
We want to know if you are willing to be in a research study. A research study is a way to find out new information about something. In this research study, we want to find out the best way to teach new words to kids.

This research study being done by Mr. Kade Johnson as part of his doctoral dissertation in educational psychology at the University of Utah. Even though Mr. Johnson works as a school psychologist in Florence Unified School District (FUSD), this research study is not sponsored by FUSD. But, FUSD and the principal at Anthem K-8 have said it is okay for Mr. Johnson to do this study at your school.

Why are we asking you to be in this research study?
We are asking you to be in this research study because we want to learn more about teaching new words to kids. We also want to know if kids who have different learning abilities learn new words in the same way.

WHAT HAPPENS IN THE RESEARCH STUDY?
If you decide to be in this research study and your parent or guardian agrees, this is what will happen. First, you, and a lot of other kids, will be given a test to find out how you learn. Then, some kids will be picked to be in the rest of the research study. You may or may not be one of the kids picked. If you are picked, you will be taught words from the Esperanto international language on a computer. You will learn a different list of words each week for three weeks. Then, we will see if you remember the words you were taught later on. At the end of each session, you will get to pick a small piece of candy, small toy, or sticker from a prize box. It will take about 10-25 minutes when we teach you the words. It will only take about one minute to see if you remember the words when we check later on. The test at the beginning that tells us how you learn will take about 15-30 minutes. Overall, the whole research study will take six weeks, and you will participate once a week for anywhere from one minute to 25 minutes. We will work with your teacher to make sure you do not miss important things in class. If you are pulled out of class during a time you do not want to miss, you can just tell your teacher and we will come get you at another time.

WILL ANY PART OF THE RESEARCH STUDY HURT YOU?
By participating in this research study, you will leave your class six times, and you might miss some things your class is doing. We will talk to your teacher to make sure we take you from class only during times when you will not miss important things.

WILL THE RESEARCH STUDY HELP YOU OR ANYONE ELSE?
By the end of this research study, you will have learned the meaning of and how to say 18 Esperanto words. Also, being in this study will help us to better understand ways to teach new words to other kids.

WHO WILL SEE THE INFORMATION ABOUT YOU?
All of your records about this research study will be kept in a locked file cabinet and on a computer that uses a password so no one but the researcher can look at them. We will not tell anyone else that you are in the study.

If you tell us that you want to hurt yourself, we will tell other adults about it so that we can help you feel better.

WHAT IF YOU HAVE ANY QUESTIONS ABOUT THE RESEARCH STUDY?
It is okay to ask questions. If you don't understand something, you can ask us. We want you to ask questions now and anytime you think of them. If you have a question later that you didn't think of now, you can ask Mr. Kade Johnson anytime while you are at school, or ask your parents to call him at 520-723-6400. You or your parents can also talk to Dr. Amy Fuller, Florence Unified School District Assistant Superintendent, if you have any questions or concerns by calling (520) 866-3512 or emailing afuller@fusdaz.org.

DO YOU HAVE TO BE IN THE RESEARCH STUDY?
You do not have to be in this study if you don't want to. Being in this study is up to you. No one will be upset if you don't want to do it. Even if you say yes now, you can change your mind later and tell us you want to stop. You can take your time to decide. You can talk to your parent or guardian before you decide. We will also ask your parent or guardian to give their permission for you to be in this study. But even if your parent or guardian say "yes" you can still decide not to be in the research study.

Agreeing to be in the study
I was able to ask questions about this study. Signing my name at the bottom means that I agree to be in this study. My parent or guardian and I will be given a copy of this form after I have signed it.

_____
Printed Name

_____ _____
Sign your name on this line Date

_____
Printed Name of Person Obtaining Assent

_____ _____
Signature of Person Obtaining Assent Date

The following should be completed by the study member conducting the assent process if the participant agrees to be in the study. Initial the appropriate selection:

_____ ***The participant is capable of reading the assent form and has signed above as documentation of assent to take part in this study.***

_____ ***The participant is not capable of reading the assent form, but the information was verbally explained to him/her. The participant signed above as documentation of assent to take part in this study.***

APPENDIX E

ESPERANTO WORDS

Unknown Esperanto Words

Each Esperanto word is a noun containing exactly 5 letters and 2 syllables.

Set One

Esperanto Word	English Translation
balen	whale
pokal	cup
vaker	cowboy
kuler	spoon
capel	hat
limak	snail

Set Two

Esperanto Word	English Translation
glavo	sword
kasko	helmet
drato	wire
blato	cockroach
hundo	dog
kanap	couch

Set Three

Esperanto Word	English Translation
korbo	basket
breto	shelf
teler	plate
cervo	deer
kusen	pillow
pulvi	rain

APPENDIX F

KNOWN WORDS

Known Dolch Words

Each known word is a noun containing exactly 4 letters and 1 syllable.

Pool of 35 Known Words	Known Words Used During High OTR Condition
Baby	Home
Back	Corn
Ball	Door
Bear	Fish
Bell	Boat
Bird	Baby
Boat	Time
Cake	Back
Coat	
Corn	
Doll	Known Words Used During Medium OTR Condition
Door	Hand
Duck	Bird
Farm	Snow
Feet	Wood
Fish	Feet
Game	Name
Girl	
Hand	
Head	Known Words Used During Low OTR Condition
Hill	Duck
Home	Girl
Milk	Game
Name	Tree
Nest	
Rain	
Ring	
Seed	
Shoe	
Snow	
Song	
Time	
Tree	
Wind	
Wood	

APPENDIX G

DATA COLLECTION SHEET

Data Collection Sheet – Set 1

Name: _____ Participant #: _____

Instructional Condition (check one): ☐ # 1 – 44 OTR, 8 Known Words – Blue
☐ # 2 – 27 OTR, 6 Known Words – Yellow
☐ # 3 – 14 OTR, 4 Known Words – Red

Session Dates:
Instruction: _____ 1-Week Retention: _____ 3-Week Retention: _____

Total instructional minutes (# of seconds / 60 = total minutes): _____

Esperanto Words	Translation (tally errors)		Known Words
balen	whale	_____	1) _____
pokal	cup	_____	2) _____
vaker	cowboy	_____	3) _____
kuler	spoon	_____	4) _____
capel	hat	_____	5) _____
limak	snail	_____	6) _____
			7) _____
			8) _____

One Week Retention Measure
- Check if correctly pronounced and translated in 3 seconds. Record incorrect responses.

☐ balen – whale _____
☐ pokal – cup _____
☐ vaker – cowboy _____
☐ kuler – spoon _____
☐ capel – hat _____
☐ limak – snail _____
Total Correct: _____

3-week Retention Measure
- Check if correctly pronounced and translated in 3 seconds. Record incorrect responses.

☐ balen – whale _____
☐ pokal – cup _____
☐ vaker – cowboy _____
☐ kuler – spoon _____
☐ capel – hat _____
☐ limak – snail _____
Total Correct: _____

Efficiency Calculations
- Number of words learned or retained divided by the number of instructional minutes equals the number of words learned or retained per minute of instruction.
- Words learned or retained / time = efficiency.

Initial learning:
_____ / _____ = _____

1-week Retention:
_____ / _____ = _____

3-week Retention:
_____ / _____ = _____

APPENDIX H

SCRIPT AND PROCEDURAL STEPS

Script and Procedural Steps

1. Position the laptop computer on the table directly in front of the student, with the screen positioned approximately two feet from their eyes.
2. Sit next to the student, within arms length of the laptop computer.
3. Say **"I am going to show you some words on this computer screen. Some are from the Esperanto language, and the rest are English words you already know. I will teach how to say the Esperanto words, and what they mean. Listen carefully because you will need to tell me the name and definition of these words each time you see them."**
4. Start the timer and immediately present the first word in the PowerPoint presentation
5. When each Esperanto word is initially presented, tell the student the pronunciation and English translation. When a blank PowerPoint slide appears on the computer screen, this means that a new Esperanto word will be initially presented on the next slide. When this happens, advance to the next slide and say, **"This word is _____. _____ means _____."** (e.g. "This word is pilko. Pilko means ball.")
6. After initially teaching each Esperanto word, ask the student to restate the word, its English translation, and use it in a sentence by saying **"Now you say the word and tell me what it means."** After the student correctly responds say **"Good job. Now, use the word in a short sentence."** Provide prompting or assistance if the child does not immediately use the word in a sentence. After the child uses the word in a sentence, say **"Good job. Remember how to say this word, and what it means because you will need to tell me each time you see it."**
7. After each correct response, immediately advance to the next word in the PowerPoint presentation by pushing Enter or the right arrow key on the keyboard.
8. Throughout the entire intervention, provide verbal reinforcement, such as **"good job"** or **"that's right"** each time an Esperanto word is correctly responded to. Do this even when the Esperanto words are presented in the known word positions of the IR sequence.
9. Throughout the entire intervention, if the student responds incorrectly or fails to respond within 3 seconds, provide corrective feedback by saying **"This word is _____. _____ means _____."** (e.g. "This word is pilko. Pilko means ball."). Do this even when the Esperanto words are presented in the known word positions of the IR sequence. Note any errors on the data collection sheet.
10. After corrective feedback is provided, prompt students to correctly respond to the Esperanto word before moving on by saying "**Now you say the word and tell me what it means."** Provide verbal reinforcement such as **"good job"** or **"that's right"** for correct responses after corrective feedback is provided. Do this even when the Esperanto words are presented in the known word positions of the IR sequence.
11. When the known Dolch words are presented, if the student does not independently respond, prompt them by saying, **"Tell me the word."** If a known word is incorrectly stated, provide corrective feedback by saying **"This word is _____."** and prompt the student to restate the word. No verbal reinforcement is necessary when students respond to known words.
12. Immediately after the student responds to the last word in the PowerPoint presentation, stop the timer and record the total session length on the data collection sheet.
13. At the end of the session, offer the student a prize from the prize box.

APPENDIX I

SESSIONS

Sessions

This table outlines the content and sequence of events (left to right) during each of the six sessions.

Session Number	1-week Retention Measure	3-week Retention Measure	Instructional Session
1			High OTR Condition
2	High OTR Condition		Med. OTR Condition
3	Med. OTR Condition		Low OTR Condition
4	Low OTR Condition	High OTR Condition	
5		Med. OTR Condition	
6		Low OTR Condition	

APPENDIX J

INSTRUCTIONAL CONDITION AND ESPERANTO

SET COMBINATIONS

Participant ID Number for below average IQ group

		1	2	3	4	5	6	7	8	9	10	11	12
Condition 1 (8 Known - 44 OTR)	Esperanto Set 1	■			■			■			■		
	Esperanto Set 2		■			■			■			■	
	Esperanto Set 3			■			■			■			■
Condition 2 (6 Known - 27 OTR)	Esperanto Set 1			■			■			■			■
	Esperanto Set 2	■			■			■			■		
	Esperanto Set 3		■			■			■			■	
Condition 3 (4 Known - 14 OTR)	Esperanto Set 1		■			■			■			■	
	Esperanto Set 2			■			■			■			■
	Esperanto Set 3	■			■			■			■		

Participant ID Number for average IQ group

		13	14	15	16	17	18	19	20	21	22	23	24
Condition 1 (8 Known - 44 OTR)	Esperanto Set 1	■			■			■			■		
	Esperanto Set 2		■			■			■			■	
	Esperanto Set 3			■			■			■			■
Condition 2 (6 Known - 27 OTR)	Esperanto Set 1			■			■			■			■
	Esperanto Set 2	■			■			■			■		
	Esperanto Set 3		■			■			■			■	
Condition 3 (4 Known - 14 OTR)	Esperanto Set 1		■			■			■			■	
	Esperanto Set 2			■			■			■			■
	Esperanto Set 3	■			■			■			■		

APPENDIX K

TREATMENT ACCEPTABILITY FORM

Treatment Acceptability Form

Name: _____ Participant #: _____

Date: _____

After completing the final retention measurement during week six, ask the following questions to each participant and record their responses.

1. Do you feel like the computer program helped you learn new words from the Esperanto language?

   YES NO

2. Did you enjoy learning the new words from the Esperanto language?

   YES NO

3. Did you like the session with the blue, yellow, or red computer screen the best?

   BLUE YELLOW RED

4. Why did you like that color session the best?

   _____
   _____
   _____
   _____

5. Why didn't you like the other color sessions as much?

   _____
   _____
   _____
   _____

APPENDIX L

TREATMENT INTEGRITY CHECKLIST

Treatment Integrity Checklist

Name: _____ Participant #: _____

Instructional Condition (check one):
- ☐ # 1 – 44 OTR, 8 Known Words – Blue
- ☐ # 2 – 27 OTR, 6 Known Words – Yellow
- ☐ # 3 – 14 OTR, 4 Known Words – Red

Date of instructional session: _____

Check the box below if procedure was followed:

☐ 1. The laptop computer was positioned on a table directly in front of the student, with the screen positioned approximately two feet from their eyes.

☐ 2. The experimenter was sitting next to the student, within arms length of the laptop computer.

☐ 3. The experimenter said **"I am going to show you some words on this computer screen. Some are from the Esperanto language, and the rest are English words you already know. I will teach how to say the Esperanto words, and what they mean. Listen carefully, because you will need to tell me the name and definition of these words each time you see them."**

☐ 4. The experimenter started the timer and immediately presented the first word in the PowerPoint presentation.

☐ 5. The experimenter stated the pronunciation and English translation for each of the six Esperanto words when they were initially presented by saying **"This word is _____. _____ means _____."** (e.g. "This word is pilko. Pilko means ball."). (Check for each of the 6 Esperanto words: ☐ ☐ ☐ ☐ ☐ ☐)

☐ 6. The experimenter asked the student to restate each Esperanto word, its English translation, and use it in a sentence when it was initially presented by saying **"Now you say the word and tell me what it means."** After the student correctly responded the experimenter said **"Good job. Now, use the word in a short sentence."** (Check for each of the 6 Esperanto words: ☐ ☐ ☐ ☐ ☐ ☐)

☐ 7. After the child initially restated and used each Esperanto word in a sentence, the experimenter said **"Good job. Remember how to say this word, and what it means because you will need to tell me each time you see it."** (Check for each of the 6 Esperanto words: ☐ ☐ ☐ ☐ ☐ ☐)

☐ 8. The experimenter immediately pushed Enter or the right arrow key on the keyboard after each correct response to advance to the next word in the PowerPoint presentation.

☐ 9. The experimenter provided verbal reinforcement, such as **"good job"** or **"that's right"** when Esperanto words were correctly stated and defined.

☐ 10. The experimenter provided corrective feedback if the student responded incorrectly or failed to respond within 3 seconds by saying **"This word is _____. _____ means _____."** (e.g. "This word is pilko. Pilko means ball."). Any errors were noted on the data collection sheet. (If corrective feedback was not necessary, check this box as being followed.)

☐ 11. If corrective feedback was provided, the experimenter said **"Now you say the word and tell me what it means."** Verbal reinforcement such as **"good job"** or "**that's right"** was provided for correct responses after corrective feedback was provided. (If corrective feedback was not necessary, check this box as being followed.)

☐ 12. The experimenter stopped the timer after the student responded to the last word on the PowerPoint presentation, and recorded the total session length on the data collection sheet.

☐ 13. The experimenter offered the student a prize from the prize box at the end of the session.

Total steps followed: _____

Percent of steps followed: _____
(Total steps followed / 13 = % of steps followed)

APPENDIX M

INTERSCORER AGREEMENT FORM

Interscorer Agreement Form

Name: _____ Participant #: _____

Instructional Condition (check one): ☐ # 1 – 44 OTR, 8 Known Words – Blue
☐ # 2 – 27 OTR, 6 Known Words – Yellow
☐ # 3 – 14 OTR, 4 Known Words – Red

☐ 1-week retention measure (Date: _____)

☐ 3-week retention measure (Date: _____)

Retention Measure – Set 1
- Check if correctly pronounced and translated in 3 seconds. Record any incorrect responses.

☐ balen – whale _____
☐ pokal – cup _____
☐ vaker – cowboy _____
☐ kuler – spoon _____
☐ capel – hat _____
☐ limak – snail _____
Total Correct: _____

Efficiency Calculation
- Number of words retained divided by the number of instructional minutes equals the number of words retained per minute of instruction, or efficiency.
- Words retained / time = efficiency
- Transfer the number of instructional minutes from the data collection sheet.

_____ / _____ = _____

Interscorer Agreement for Word Retention
- Number of score agreements divided by the by the total number of words presented during the retention measure (6) times 100.
- (agreements / 6)100 = interscorer agreement

(_____ / 6)100 = _____

Interscorer Agreement for Efficiency
Do the efficiency calculations match? (circle one)

YES NO

REFERENCES

Albers, A. E., & Greer, R. D. (1991). Is the three-term contingency trial a predictor of effective instruction? *Journal of Behavioral Education*, *1*(3), 337-354.

Allor, J. H., Champlin, T. M., Gifford, D. B., & Mathes, P. G. (2010). Methods for increasing the intensity of reading instruction for students with intellectual disabilities. *Education and Training in Autism and Developmental Disabilities*, *45*(4), 500-511.

Betts, E. A. (1946). *Foundations of reading instruction*. New York: American Book.

Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Delzell, L., & Algozzine, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children*, *72*, 392-408.

Browder, D. M., & Shear, S. M. (1996). Interspersal of known items in a treatment package to teach sight words to students with behavior disorders. *Journal of Special Education*, *29*, 400-413.

Browder, D. M., & Xin, Y. (1998). A meta-analysis and review of sight word research and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education*, *32*(3), 130-153.

Bunn, R., Burns, M. K., Hoffman, H. H., & Newman, C. L. (2005). Using incremental rehearsal to teach letter identification with a preschool-aged child. *Journal of Evidence Based Practice for Schools*, *6*, 124-134.

Burns, M. K. (2004a). Age as a predictor of acquisition rates as measured by curriculum-based assessment: Evidence of consistency with cognitive research. *Assessment for Effective Intervention*, *29*(2), 31-38.

Burns, M. K. (2004b). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education*, *25*(3), 167-173.

Burns, M. K. (2005). Using incremental rehearsal to practice multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children*, *28*, 237-249.

Burns, M. K. (2007a). Comparison of drill ratio and opportunities to respond when rehearsing sight words with a child with mental retardation. *School Psychology Quarterly, 22,* 250-263.

Burns, M. K. (2007b). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School Psychology Quarterly*, *22*, 297−313.

Burns, M. K., & Boice, C. H. (2009). Comparison of the relationship between words retained and intelligence for three instructional strategies among students with low IQ. *School Psychology Review*, *38*, 284-292.

Burns, M. K., Dean, J. D., & Foley, S. (2004). Preteaching unknown key words with incremental rehearsal to improve reading fluency comprehension with children identified as reading disabled. *Journal of School Psychology*, *42*, 303-314.

Burns, M. K., & Kimosh, A. (2005). Using incremental rehearsal to teach sight words to adult students with moderate mental retardation. *Journal of Evidence-Based Practices for Schools*, *6*, 137-147.

Burns, M. K., & Sterling-Turner, H. E. (2010). Comparison of efficiency measures for academic interventions based on acquisition and maintenance. *Psychology in the Schools*, *47*(2), 126-134.

Burns, M. K., Tucker, J. A., Hauser, A., Thelen, R. L., Holmes, K. J., & White, K. (2002). Minimum reading fluency rate necessary for comprehension: A potential criterion for curriculum-based assessments. *Assessment for Effective Intervention*, *28*(1), 1-7.

Cates, G. L., & Skinner, C. H. (2000). Getting remedial mathematics students to prefer homework with 20% and 40% more problems: An investigation of the strength of the interspersing procedure. *Psychology in the Schools*, *37*, 339-347.

Cates, G. L., Skinner, C. H., Waison, T. S., Meadows, T. Y., Weaver, A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based decision-making: An evaluation of interspersing procedures. *School Psychology Review*, *32*, 601-616.

Cates, G. L., Skinner, C. H., Watkins, C. E., Rhymer, K. N., McNeill, B. S., & McCurdy, M. (1999). Effects of interspersing additional brief math problems on student performance and perception of math assignments: Getting students to prefer to do more work. *Journal of Behavioral Education*, *9*, 177-193.

Chase, D. H., & Symonds, P. M. (1992). Practice vs. motivation. *Journal of Educational Psychology*, *84*, 282-289.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conners, F. A. (1992). Reading instruction for students with moderate mental retardation: Review and analysis of research. *American Journal of Mental Retardation*, *96*, 577-597.

Coding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension of using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioral Education*, *19*, 93-105.

Cooke, N. L., Guzaukas, R., Pressley, L. S., & Kerr, K. (1993). Effects of using a ratio of new items to review items during drill and practice: Three experiments. *Education and Treatment of Children*, *16*, 213-234.

Cuvo, A. J., Davis, P. K., & Gluck, T. C. (1991). Cumulative and interspersal task sequencing in self-paced training for persons with mild handicaps. *Mental Retardation*, *29*, 335-342.

Daly, E. J., III, Hintze, J. M., & Hamler, K. R. (2000). Improving practice by taking steps toward technological improvements in academic interventions in the new millennium. *Psychology in the Schools*, *37*, 61-72.

Daly, E. J., III., & McCurdy, M. (2002). Getting it right so they can get it right: An overview of the special series. *School Psychology Review*, *31*, 453-458.

Dehaene, S. & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 314-326.

Dempster, F. N. (1991). Synthesis of research on reviews and tests. *Educational Leadership*, *48*, 71-76.

Dolch, E. W. (1936). A basic sight vocabulary. *Elementary School Journal*, *36*, 456-460.

Dunlap, G. (1984). The influence of task variation and maintenance tasks on the learning and affect of autistic children. *Journal of Experimental Child Psychology*, *37*, 41-64.

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis*, *13*, 619-627.

Dunn, L. M., & Dunn, L. M. (1981). *Manual for forms L and M of the Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.

Ebbinghaus, H. (1885). Über das gedächtnis. Leipzig: Duncker & Humbolt (*Memory*; H. A. Ruger & C. E. Bussenius, Trans. New York: Teachers College, 1913; reprinted by Dover, 1964.)

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167-188.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* 39, 175-191.

Fry, E. B., & Kress, J. E. (2006). *The reading teacher's book of lists, 5th Ed*. San Francisco: Jossey-Bass.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239-256.

Gates, A. I. (1930). *Interest and ability in reading*. New York: Macmillan.

Gickling, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprehension. *Journal of Learning Disabilities*, *11*, 559-566.

Gickling, E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (3rd ed., pp. 587–595). Washington, DC: National Association of School Psychologists.

Gickling, E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children*, *52*, 205-218.

Glenberg, A. M., & Smith, S. M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior*, *20*, 110-119.

Gravois, T. A., & Gickling, E. E. (2008). Best practices in instructional assessment. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 503-518). Bethesda, MD: National Association of School Psychologists.

Greenwood, C. R., Delquadri, J., & Hall, R. V. (1984). Opportunity to respond and student academic performance. In W. Heward, T. Heron, D. Hill & J. Trap-Porter (Eds.), *Behavior analysis in education* (pp. 58-88). Columbus, OH: Charles E. Merrill.

Hargis, C. H., Terhaar-Yonkers, M., Williams, P. C., & Reed, M. T. (1988). Repetition requirements for word recognition. *Journal of Reading*, *31*, 320-327.

Haring, N. G. & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton & C. L. Hanson (Eds.), *The fourth R: Research in the classroom* (pp. 23-40). Columbus, OH: Merrill

Hawkins, J., Skinner, C. H., & Oliver, R. (2005). The effects of task demands and additive interspersal ratios on fifth-grade students' mathematics accuracy. *School Psychology Review*, *34*, 543-555.

Janton, P. (1993). *Esperanto: Language, literacy, and community*. Albany, NY: State University of New York.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of Individual Differences in Reading Comprehension and Reading Fluency. *Journal of Educational Psychology*, *95*(4), 719-729.

Jensen, A. R. (1989). The relationship between learning and intelligence. *Learning and Individual Differences*, *1*, 37-62.

Jones, D., & Christensen, C. A. (1999). Relationship between automaticity in hand writing and student's ability to generate written text. *Journal of Educational Psychology*, *91*, 44-49.

Joseph, L., & Nist, L. (2006). Comparing the effects of unknown-known ratios on word reading learning versus learning rates. *Journal of Behavioral Education*, *15*(2), 69-79.

Joseph, L.M., Schisler, R. (2007). Getting the "most bang for your buck": Comparison of the effectiveness and efficiency of phonic and whole word reading techniques during repeated reading lessons. *Journal of Applied School Psychology*, *24*(1), 69-90.

Joseph, L. M., & Seery, M. E. (2004). Where is the phonics? A review of the literature on the use of phonetic analysis with students with mental retardation. *Remedial and Special Education*, *25,* 88-94.

Kamps, D. M., Dugan, E. P., Leonard, B. R., & Daoust, P. M. (1994). Enhanced small group instruction using choral responding and student interaction for children with autism and developmental disabilities. *American Journal on Mental Retardation*, *99*, 60-73.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test, Second Edition*. Bloomington, MN: Pearson, Inc.

Kupzyk, S., Daly, E. J., III, & Andersen, M. N. (2011). A comparison of two flash-card methods for improving sight-word reading. *Journal of Applied Behavior Analysis*, *44*, 781-792.

Levy, B. A., Abello, B., & Lysynchuk, L. (1997). Transfer from word training to reading on context: Gains in reading fluency and comprehension. *Learning Disability Quarterly*, *20*, 173-188.

Logan, G. D., & Klapp, S. T. (1991). Alphabet arithmetic: Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 179-195.

MacQuarrie, L. L., Tucker, J. A., Bums, M. K., & Hartman, B. (2002). Comparison of retention rates using traditional, drill sandwich, and incremental rehearsal flash card methods. *School Psychology Review*, *31*, 584-595.

Matchett, D. L., & Burns, M. K. (2009). Increasing word recognition fluency with an English language learner. *Journal of Evidence Based Practices in Schools*, *10*, 194-209.

McCurdy, M., Skinner, C. H., Grantham, K., Watson, T., & Hindman, P. M. (2001). Increasing on-task behavior in an elementary student during mathematics seatwork by interspersing additional brief problems. *School Psychology Review*, *30*, 23-32.

Neef, N. A., Iwata, B. A., & Page, T. J. (1977). The effects of known-item interspersal on acquisition and retention of spelling and sight reading words. *Journal of Applied Behavior Analysis*, *10*, 738.

Neef, N. A., Iwata, B. A., & Page, T. J. (1980). The effects of interspersal training versus high-density reinforcement of spelling acquisition and retention. *Journal of Applied Behavior Analysis*, *13*, 153-158.

Nevin, J. A., Mandell, C., & Atak, J. R. (1983). The analysis of behavioral momentum. *Journal of the Experimental Analysis of Behavior*, *39*, 49-59.

Nist, L., & Joseph, L. M. (2008). Effectiveness and efficiency of flashcard drill instructional methods on urban first-graders' word recognition, acquisition, maintenance, and generalization. *School Psychology Review*, *37*, 294-308.

Petersen-Brown, S., & Burns, M.K. (2011). Adding a vocabulary component to incremental rehearsal to enhance retention and generalization. *School Psychology Quarterly*, *26*, 245-255.

Pummel, M. L. (2011). *The effects of the taped-problems time-delay intervention to teach sight words using an iPod* (Unpublished master's thesis). University of Utah, Salt Lake City.

Richardson, D. (1988). *Esperanto: Learning and using the international language*. El Cerrito, CA: Esperanto League for North America.

Roberts, M. L., & Shapiro, E. S. (1996). Effects of instructional ratios on students' reading performance in a regular education program. *Journal of School Psychology*, *34*, 73-91.

Roberts, M. L., Turco, T. L., & Shapiro, E. S. (1991). Differential effects of fixed instructional ratios on students' progress in reading. *Journal of Psychoeducational Assessment*, *9*, 308-318.

Robinson, S. L., & Skinner, C. H. (2002). Interspersing additional easier items to enhance mathematics performance on subtests requiring different task demands. *School Psychology Quarterly*, *17*, 191-205.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, Jerome M. Sattler Publisher Inc.

Schloss, P. J., Alper, S. K., & Young, H. (1995). Acquisition of functional sight words in community-based recreation settings. *The Journal of Special Education*, *29*, 84-96.

Schmidgall, M., & Joseph, L. (2007). Comparison of phonic analysis and whole word reading on first graders' cumulative words read and cumulative reading rate: An extension in examining instructional effectiveness and efficiency. *Psychology in the Schools*, *44*(4), 319-332.

Schnorr, J. M. (1989). Practicing math facts on the computer. *Teacher Education and Special Education*, *12*, 65-69.

Schwanenflugel, P. J., Meisinger. E. B., Kuhn, M. R., Strauss, G. P., & Morris. R. D. (2006). Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly*, *41*, 496-522.

Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research*, *64*, 253-286.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention* (4th ed.). New York: Guilford Press.

Shapiro, E. S., & Ager, C. (1992). Assessment of special education students in regular education programs: Linking assessment to instruction. *Elementary School Journal*, *92*(3), 283-296.

Skinner, C. H. (2002). An empirical analysis of interspersal research evidence, implications, and applications of discrete task completion hypothesis. *Journal of School Psychology*, *40*, 347-368.

Skinner, C. H. (2008). Theoretical and applied implications of precisely measuring learning rates. *School Psychology Review*, *37*, 309-315.

Skinner, C. H. (2010). Applied comparative effectiveness researchers must measure learning rates: A commentary on efficiency articles. *Psychology in the Schools*, *47*(2), 166-172.

Skinner, C. H., Belfiore, P. J., Mace, H. W., Williams-Wilson, S., & Johns, G. A. (1997). Altering response topography to increase response efficiency and learning rates. *School Psychology Quarterly*, *12*, 54-64.

Skinner, C. H., Belfiore, P. J., & Watson, T. S. (1995). Assessing the relative effects of interventions in students with mild disabilities: Assessing instructional time. *Assessment in Rehabilitation and Exceptionality*, *2*, 207-220.

Skinner, C. H., Fletcher, P. A., & Hennington, C. (1996). Increasing learning rates by increasing student response rates: A summary of research. *School Psychology Quarterly*, *11*, 313-325.

Skinner, C. H., Fletcher, P. A., Wildmon, M., & Belfiore, P. J. (1996). Improving assignment preference through interspersing additional problems: Brief versus easy problems. *Journal of Behavioral Education, 6*, 427-437.

Skinner, C. H., Hall-Johnson, K., Skinner, A. L., Cates, G., Weber, J., & Johns, G. (1999). Enhancing perceptions of mathematics assignments by increasing relative rates of problem completion through the interspersal technique. *Journal of Experimental Education*, *68*, 43-59.

Skinner, C. H., Robinson, S. L., Johns, G. A., Logan, P., & Belfiore, P. J. (1996). Applying Herrnstein's matching law to influence students' choice to complete difficult academic tasks. *Journal of Experimental Education*, *65*, 5-17.

Skinner, C. H., Turco, T. L., Beatty, K. L., & Rasavage, C. (1989). Cover, copy, and compare: A method for increasing multiplication performance. *School Psychology Review*, *18*, 412-420.

Sutherland, K. S., & Wehby, J. H. (2001). Exploring the relationship between increased opportunities to respond to academic requests and the academic behavioral outcomes of students with EBD: A review. *Remedial and Special Education*, *22*, 113-121.

Symonds, P. M. & Chase, D. H. (1992). Practice vs. motivation. *Journal of Educational Psychology*, *84*, 282-289.

Szadokierski, I., & Burns, M. K. (2008). Analogue evaluation of the effects of opportunities to respond and ratios of known items within drill rehearsal of Esperanto words. *Journal of School Psychology*, *46*, 593-609.

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension. *Journal of Educational Psychology*, *89*, 276-288.

Tucker, J. A. (1989). *Basic flashcard technique when vocabulary is the goal*. Unpublished teaching materials, School of Education, University of Tennessee at Chattanooga.

Volpe, R. J., Burns, M. K., DuBois, M., & Zaslofsky, A. (2011). Computer-assisted tutoring: Teaching letter sounds to kindergarten students using incremental rehearsal. *Psychology in the Schools*, *48*(4), 332-342.

Volpe, R. J., Mulé, C. M., Briesch, A. M., Joseph, L. M., & Burns, M. K. (2011). A comparison of two flashcard drill methods targeting word recognition. *Journal of Behavioral Education*, *20*(2), 117-137.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Wilson, R., Majsterek, D., & Simmons, D. (1996). The effects of computer-assisted versus teacher-directed instruction on the multiplication performance of elementary students with learning disabilities. *Journal of Learning Disabilities*, *29*, 382-390.